

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
21 February 2002 (21.02.2002)

PCT

(10) International Publication Number  
**WO 02/14495 A2**

(51) International Patent Classification<sup>7</sup>: **C12N 15/00**

21754 (US). COPELAND, Neal, G. [US/US]; 10022  
Pebble Beach Terrace, Ijamsville, MD 21754 (US).

(21) International Application Number: **PCT/US01/25507**

(22) International Filing Date: 14 August 2001 (14.08.2001)

(74) Agent: NOONAN, William, D.; Klarquist, Sparkman,  
LLP, One World Trade Center, Suite 1600, 121 S.W.  
Salmon Street, Portland, OR 97204 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/225,164 14 August 2000 (14.08.2000) US  
60/271,632 26 February 2001 (26.02.2001) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,  
ZW.

(71) Applicant (*for all designated States except US*): THE  
GOVERNMENT OF THE UNITED STATES OF  
AMERICA, as represented by THE SECRETARY,  
DEPARTMENT OF HEALTH AND HUMAN SER-  
VICES, THE NATIONAL INSTITUTES OF HEALTH  
[US/US]; Office of Technology Transfer, 6001 Executive  
Boulevard, Suite #325, Rockville, MD 20852 (US).

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): COURT, Donald,  
L. [US/US]; 502 Magnolia Avenue, Frederick, MD 21702  
(US). YU, Daiguan [US/US]; 1418 Taney Avenue, H301,  
Frederick, MD 21702 (US). LEE, E-Chiang [State-  
less/US]; 2402 Dominion Drive, #2C, Frederick, MD  
21702 (US). ELLIS, Hilary, M. [US/US]; 4713 Edgefield  
Road, Bethesda, MD 20814 (US). JENKINS, Nancy, A.  
[US/US]; 10022 Pebble Beach Terrace, Ijamsville, MD

Published:

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

WO 02/14495 A2

(54) Title: ENHANCED HOMOLOGOUS RECOMBINATION MEDIATED BY LAMBDA RECOMBINATION PROTEINS

(57) Abstract: Disclosed herein are methods for generating recombinant DNA molecules in cells using homologous recombination mediated by recombinases and similar proteins. The methods promote high efficiency homologous recombination in bacterial cells, and in eukaryotic cells such as mammalian cells. The methods are useful for cloning, the generation of transgenic and knockout animals, and gene replacement. The methods are also useful for subcloning large DNA fragments without the need for restriction enzymes. The methods are also useful for repairing single or multiple base mutations to wild type or creating specific mutations in the genome. Also disclosed are bacterial strains which are useful for high-efficiency homologous recombination.

## ENHANCED HOMOLOGOUS RECOMBINATION MEDIATED BY LAMBDA RECOMBINATION PROTEINS

### FIELD

5       The present disclosure relates to methods to enhance homologous recombination in bacteria and eukaryotic cells using recombination proteins, such as those derived from bacteriophage lambda. It also relates to methods for modifying genomic DNA in bacterial artificial chromosomes (BACs) and to subcloning of genomic DNA from BACs into multicopy plasmids.

10

### BACKGROUND OF THE INVENTION

Concerted use of restriction endonucleases and DNA ligases allows *in vitro* recombination of DNA sequences. The recombinant DNA generated by restriction and ligation may be amplified in an appropriate microorganism such as *E. coli*, and used for  
15   diverse purposes including gene therapy. However, the restriction-ligation approach has two practical limitations: first, DNA molecules can be precisely combined only if convenient restriction sites are available; second, because useful restriction sites often repeat in a long stretch of DNA, the size of DNA fragments that can be manipulated are limited, usually to less than about 25 kilobases.

20       Homologous recombination, generally defined as an exchange between homologous segments anywhere along a length of two DNA molecules, provides an alternative method for engineering DNA. In generating recombinant DNA with homologous recombination, a microorganism such as *E. coli*, or a eukaryotic cell such as a yeast or vertebrate cell, is transformed with exogenous DNA. The center of the  
25   exogenous DNA contains the desired transgene, whereas each flank contains a segment of homology with the cell's DNA. The exogenous DNA is introduced into the cell with standard techniques such as electroporation or calcium phosphate-mediated

transfection, and recombines into the cell's DNA, for example with the assistance of recombination-promoting proteins in the cell.

In generating recombinant DNA by homologous recombination, it is often advantageous to work with short linear segments of DNA. For example, a mutation  
5 may be introduced into a linear segment of DNA using polymerase chain reaction (PCR) techniques. Under proper circumstances, the mutation may then be introduced into cellular DNA by homologous recombination. Such short linear DNA segments can transform yeast, but subsequent manipulation of recombinant DNA in yeast is laborious. It is generally easier to work in bacteria, but linear DNA fragments do not  
10 readily transform bacteria (due in part to degradation by bacterial exonucleases). Accordingly, recombinants are rare, require special poorly-growing strains (such as RecBCD- mutant strains) and generally require thousands of base pairs of homology. Thus, improved methods of promoting homologous recombination in bacteria are needed.

15 In eukaryotic cells, targeted homologous recombination provides a basis for targeting and altering essentially any desired sequence in a duplex DNA molecule, such as targeting a DNA sequence in a chromosome for replacement by another sequence. The approach may be useful for treating human genetic diseases.

Homologous recombination has been used to create knock-out mutants and  
20 transgenic animals, and thereby has played a critically important role in understanding gene function. Transgenic animals are organisms that contain stably integrated copies of genes or gene constructs derived from another species in the chromosome of the transgenic animal. These animals can be generated by introducing cloned DNA constructs of the foreign genes into totipotent cells by a variety of methods, including  
25 homologous recombination.

Currently methods for producing transgenics have been performed on totipotent embryonic stem cells (ES) and with fertilized zygotes. ES cells have an advantage in that large numbers of cells can be manipulated in vitro before they are used to generate transgenics. Alternatively, DNA can also be introduced into fertilized oocytes by

micro-injection into pronuclei, or injected into the germline of organisms including *C. elegans* or *Drosophila* species.

Several methods have been developed to detect and/or select for targeted site-specific recombinants between vector DNA and the target homologous chromosomal sequence (Capecchi, Science 244: 1288, 1989). Cells that exhibit a specific phenotype after recombination, such as occurs with alteration of the hypoxanthine phosphoribosyl transferase (*hprt*) gene, can be obtained by direct selection on the appropriate growth medium. Alternatively, a selectable marker such as neomycin resistance can be incorporated into a vector under promoter control, and successful transfection can be scored by selecting G418-resistant cells (Joyner et al., Nature 338: 153, 1989). Numerous other selection procedures have been described (Jasin and Berg, Genes and Development 2: 1353, 1988; Doetschman et al. Proc. Natl. Acad. Sci. (U.S.A.) 85: 8583, 1988; Dorini et al., Science 243: 1357, 1989; Itzhaki and Porter, Nucl. Acids Res. 19: 3835, 1991). Unfortunately, exogenous sequences transferred into eukaryotic cells undergo homologous recombination only at very low frequencies, even when very long homology regions are present (Koller et al. Proc. Natl. Acad. Sci. U.S.A., 88: 10730, 1991, and Snouwaert et al. Science 257: 1083, 1992). Thus, large numbers of cells must be transfected, selected, and screened in order to generate a correctly targeted homologous recombinant.

## SUMMARY OF THE DISCLOSURE

The present disclosure provides methods for cloning DNA molecules in cells having DNA encoding lambda recombinases operably linked to a de-repressible promoter, for example the lambda PL promoter. The PL promoter is activated, by temperature shift, thereby leading to expression of lambda recombinases. The lambda recombinases promote homologous recombination between nucleic acids in the cell. The nucleic acids undergoing recombination may be intrachromosomal, or may be extrachromosomal, for example in a bacterial artificial chromosome.

The present disclosure also provides methods for inducing homologous recombination using single-stranded DNA molecules, by introducing into the cell DNA capable of undergoing homologous recombination, and a single-stranded DNA binding polypeptide capable of promoting homologous recombination. Such single-stranded DNA binding polypeptides include lambda Beta, RecT, P22 Erf, and Rad52, as well as functional fragments and variants of single-stranded DNA binding polypeptides.

The present disclosure also provides bacterial cells that promote efficient homologous recombination. These bacterial cells contain one or more genes or promoters from a defective lambda prophage within the bacterial chromosome.

The disclosure also provides methods for altering eukaryotic genes by expressing recombinases operably linked to a de-repressible promoter in bacterial cells. Eukaryotic genes thus modified can be used to modify eukaryotic cells, for example to generate transgenic or knockout animals.

### BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 is a linear depiction of the defective lambda prophage as it is integrated on the *E. coli* chromosome. Prophage genes are indicated by the solid line, and *E. coli* genes by broken lines. The coding regions for the lambda recombination genes *exo*, *bet* and *gam* are approximately in the center of the defective prophage. Lambda genes *cro* through *attR*, deleted from the defective prophage, are enclosed within parentheses together with the *E. coli bioA* gene to indicate their deletion.

Fig. 2A is a schematic diagram showing classical recombinant technology, and Fig. 2B is a schematic diagram showing "recombineering" using homologous recombination, as disclosed herein.

Fig. 3 is a graph that shows the effect of induction time of the lambda proteins Beta, Exo, and Gam on recombination efficiency. A linear chloramphenicol resistance cassette was used to target prophage genes, and was electroporated into cells after the cells had been heated to 42 °C for the time indicated. Time of induction (temperature

shift to 42 °C) is plotted on the x-axis, against number of chloramphenicol-resistant recombinants obtained on the y-axis.

Fig. 4 is a graph that shows the effect of linear DNA amount on recombination. The *E. Coli* strain DY330 was induced at 42 °C for 15 minutes, and DNA encoding a linear chloramphenicol resistance cassette (<cat>) at the indicated concentration was electroporated into cells. DNA concentration encoding <cat> is plotted on the x-axis, against number of chloramphenicol-resistant recombinants obtained on the y-axis.

Fig. 5 is a graph which shows the effect of homologous arm length on recombination in *E. coli* cells induced for expression of lambda recombination proteins Beta, Exo, and Gam. A chloramphenicol resistance cassette was synthesized with homologous arms of the indicated length at its 5' and 3' end. Homologous arm length was varied from 0 to 1,000 base pairs. Length of homologous arms is plotted on the x-axis, against number of chloramphenicol-resistant recombinants (log scale) on the y-axis.

Fig. 6 is a linear depiction of the modified defective lambda prophage as integrated on the *E. coli* chromosome of DY380, EL250, and EL350 cells. This figure illustrates that the defective prophages used for BAC engineering contain the  $\lambda$  genes from *cI857* to *int*.  $P_L$  and  $P_R$  denote the lambda left and right promoters, respectively. The *gam* and *red* genes, *exo* and *bet* are under the control of  $P_L$ , which is repressed by the temperature-sensitive repressor, *cI857* at 32°C and derepressed at 42°C. *tet* replaces the segment from *cro-bioA* in DY380 cells. The *araC-P<sub>BAD</sub>flpe* cassette or the *araC-P<sub>BAD</sub>cre* cassette replaces the segment from *cro-bioA* in EL250 or EL350 cells, respectively. The promoter of the *araBAD* operon ( $P_{BAD}$ ), which can be induced by L-arabinose, controls the expression of the *flpe* or *cre* genes. Thick black lines designate the prophage while thin lines represent *E. coli* sequence. < > defines the ends of the *cro-bioA* region that was replaced with *tet*, *araC-P<sub>BAD</sub>flpe*, or *araC-P<sub>BAD</sub>cre*.

Fig. 7 illustrates a strategy for BAC engineering. This figure illustrates the relative position of the *Eno2* gene in the fully sequenced 250-kbp BAC, 284H12 and the

different steps used to introduce Cre into the last exon of *Eno2*. In the targeting cassette, *FRT* sites are denoted by ellipses, the *kan* gene by a red rectangle and the *GFPcre* fusion gene by a blue rectangle. The green boxes represent *Eno2* exons.

Fig. 8 is a schematic diagram illustrating the use of the gap repair to subclone fragments as large as 80 kbp from BACs. Fig. 8A shows as short thick black arrows the location of the 5' homologies on the amplification primers used to amplify pBR322 for subcloning by gap repair. Each primer also contains 20 nt segments at its 3' end to prime pBR322. *NotI* and *SalI* cleavage sites were included in these primers to facilitate release of the subcloned fragments from the plasmid backbone. The location of *SpeI* restriction sites near *Eno2* is also shown ("S"). *SpeI* restriction sites are not present on the linear amplified pBR322 vector. Fig. 8B shows an intermediate step in gap repair, pairing between a typical amplified pBR322 targeting cassette and the modified *Eno2* BAC. Ap, *amp* resistance gene; ORI, origin of replication.

Fig. 9 is a schematic diagram of a defective  $\lambda$  prophage. The defective prophage DY380 expressing the  $\lambda$  Exo, Beta, and Gam functions is shown with the genes under  $P_L$  promoter control and the temperature sensitive repressor, CI857. Advantages and disadvantages of the systems are described. The genes encoding Cre and Flpe are present on other derivatives of DY380 (EL250 and EL350) and replace the *tet* gene as shown.

Fig. 10 is a schematic diagram of *in vivo* cloning by recombination using gap repair of a linear vector plasmid. The method of *in vivo* cloning uses two linear DNAs, a vector and a target DNA, that have homologies to each other at their ends. Both are electroporated into competent cells to allow recombination and gap repair of the plasmid. The linear vector is made in a similar way to that shown in Fig. 8.

Fig. 11 is a schematic diagram of *in vivo* retrieving of DNA from BAC clones. Retrieving of segments up to 80 kbp from BACs into PCR-amplified vectors has been possible using recombineering techniques disclosed herein. Here only the plasmid is linearized and transformed into a recombination competent cell containing the BAC. Recombination occurs between homologies on the end of the linear vector and the



BAC. This method eliminates standard cloning technology from the BAC, and importantly, the cloned segment is never replicated *in vitro*, thereby reducing the chance of extraneous changes in the sequence.

Fig. 12 is a schematic diagram of a mini-lambda DNA circle. This lambda DNA element is not a plasmid and lacks any replication origin activity. It does contain the lambda *cI857* repressor and the *pL* operon that the repressor controls. It also contains a cassette encoding a drug marker, in this case the *tet* genes. This DNA when transformed into most strains including the BAC strains makes Int protein allowing integration of the circular DNA at the  $\lambda$  attachment site on the bacterial chromosome, and *cI857* repressor to allow repression of *pL*. The integrated mini-lambda is stable but able to be induced at 42°C to activate Gam, Beta, and Exo expression to make the cell recombination competent.

Fig. 13 is a schematic diagram of recombination of ssDNA into the genome. When ssDNA is electroporated into cells, it is bound by Beta protein and recombined into the genome or into a BAC plasmid by homology. Evidence suggests that Beta-bound ssDNA anneals to its ssDNA complement at the replication fork. The strand of DNA corresponding to that made by lagging strand synthesis is most recombinogenic suggesting that Beta simply anneals the ssDNA to a gap caused by DNA replication.

## DETAILED DESCRIPTION OF SPECIFIC EMBODIMENTS

There exists a need in the art for methods of precisely and efficiently altering predetermined endogenous genetic sequences by homologous recombination *in vivo*. There independently exists a need in the art for high-efficiency gene targeting, so as to avoid complex *in vitro* or *in vivo* selection protocols.

One or more of the foregoing needs are met by the methods disclosed herein, which include methods for cloning DNA molecules in cells using homologous recombination mediated by lambda recombinases and similar proteins.

One such method uses a cell having DNA encoding functional Exo, Beta, and Gam or fragments or variants thereof, operably linked to the a de-repressible promoter

(e.g. the PL promoter). De-repression of the de-repressible promoter (e.g. the induction of transcription from the pL promoter by inactivation of cI) induces expression of *exo*, *bet* and *gam*, and in some embodiments may be selectively activated for this purpose.

A nucleic acid (such as a polynucleotide which is homologous to a target DNA sequence) capable of undergoing homologous recombination is introduced into the cell, and cells in which homologous recombination has occurred are either selected or found by direct screening of cells. In particular embodiments, the nucleic acid introduced into the cell may be double strand DNA, or DNA with 5' overhangs.

In additional particular embodiments, at least in 1 in 5000 cells contain DNA in which homologous recombination has occurred. In further embodiments, at least 1 in 1,000 cells, or 1 in 500 cells, 1 in 100 cells, or 1 in 20 cells contain DNA in which homologous recombination has occurred.

The cell may be a eukaryotic cell, or a prokaryotic cell, such as a bacterial cell, for example an *E. coli* strain. The DNA encoding the lambda recombination proteins and PL promoter may be intrachromosomal or extrachromosomal. Similarly, the target DNA sequence may be intrachromosomal or extrachromosomal; for example, the target DNA sequence may be found in a chromosome of the cell or a plasmid (including derivatives of colE1, pSC101, p15A and shuttle vectors that replicate in both bacteria and eucaryotic cells), bacterial artificial chromosome, P1 artificial chromosome, yeast artificial chromosome, cosmid or the like.

In additional particular embodiments, the nucleic acid introduced into the cell may be double-stranded DNA or DNA with a 5' overhang, and may include a positive or negative selectable marker. The introduced nucleic acid may alter the function of a nucleic acid sequence such as a gene in the cell, or add a gene to the DNA of the cell. The cell may be treated to enhance macromolecular uptake, for example using electroporation, calcium phosphate-DNA coprecipitation, liposome mediated transfer, or other suitable methods. In other particular embodiments, the method may produce homologous recombination that alters the function of a gene in the cell, or adds a gene to the cell.

In further particular embodiments, the cell may be treated to enhance macromolecular uptake, for example with electroporation, calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, microinjection, liposome fusion, lipofection, protoplast fusion, inactivated adenovirus-mediated transfer, HVJ-liposome mediated transfer, and biolistics.

Another such method that meets one or more of the above-mentioned needs includes methods which introduce into the cell DNA capable of undergoing homologous recombination and a ssDNA binding polypeptide capable of promoting homologous recombination. In particular embodiments, the DNA is ssDNA or DNA having 3' overhangs.

The single stranded DNA (ssDNA) binding polypeptide is a type of DNA binding polypeptide which mediates double strand break repair homologous recombination by binding to ssDNA or a 3' overhang in dsDNA and promoting recombination. It promotes recombination by annealing the bound ssDNA to its complement in the cell. Examples of such ssDNA binding polypeptide include lambda Beta, *E. coli* RecT, Erf of bacteriophage P22, and Rad52. The ssDNA binding polypeptide may be introduced as a nucleic acid. For example, a nucleic acid that expresses the ssDNA binding polypeptide is introduced into a cell, such as a eukaryotic cell. Expression of the ssDNA binding polypeptide from a nucleic acid may be induced, for example, by activation of an inducible promoter. In particular embodiments, the nucleic acid may further include lambda *exo* and *gam*, and the inducible promoter may be the lambda PL promoter. In other embodiments, the ssDNA binding polypeptide is introduced into the cell as a polypeptide.

The cell used in methods disclosed herein may be a bacterial cell such as an *E. coli* strain, or a eukaryotic cell such as a mammalian cell, a stem cell, or virtually any other eukaryotic cell type. The DNA used in the method may be a single oligonucleotide sequence, or may be two or more overlapping sequences, for example with more than 10, or more than 20 base pairs of complementary overlap at either their 5' or 3' termini (in specific examples of the 5' case, the nucleic acid includes *exo* and *bet*

nucleic acid sequences). The DNA may comprise a selectable marker, and homologous recombination with the ssDNA may confer a selectable phenotype upon the cell. In particular embodiments, the cell may be treated to enhance macromolecular uptake, such as with electroporation, calcium phosphate-DNA coprecipitation, liposome mediated transfer, or other suitable methods. The effect of homologous recombination may be to alter the function of a gene in the cell, or add a gene to the cell.

In particular examples, the ssDNA is used in an amount of about 0.01  $\mu\text{M}$  to about 10 mM; or from about 0.1  $\mu\text{M}$  to about 1 mM; or from about 1  $\mu\text{M}$  to about 100  $\mu\text{M}$ . In other examples, the ssDNA binding polypeptide is used in an amount of 0.001  $\mu\text{M}$  to about 0.01  $\mu\text{M}$ , or from about 0.01  $\mu\text{M}$  to about 10 mM; or from about 0.1  $\mu\text{M}$  to about 1 mM; or from about 1  $\mu\text{M}$  to about 100  $\mu\text{M}$ .

Also disclosed are bacterial cells that may be useful in practicing the disclosed methods. These include bacterial cells harboring a defective lambda prophage of genotype  $\lambda\text{cI857 } \Delta(\text{cro-bioA})$ . In particular examples, the bacterial cells may have a selectable marker, such as an antibiotic resistance marker, upstream of the cI857 gene. In some particular examples, the disclosed bacterial cells include an inducible promoter upstream of the cI857 gene, which may be operably connected to a gene encoding a recombinase, such as flp, flpe, or cre, or a gene encoding functional fragments or variants of these recombinases. In other particular examples, the bacterial cells may contain a bacterial artificial chromosome, which may have a selectable marker, loxp, and/or FRT sites. In particular embodiments, the selectable marker on the bacterial artificial chromosome is excisable by a recombinase. In other particular examples, the bacterial artificial chromosome may have at least one exon or at least one intron of a mammalian gene.

The disclosure also includes methods of altering eukaryotic genes by expressing in a bacterial cell an intrachromosomal gene encoding a recombinase operably linked to a PL promoter. The bacterial cell also includes an extrachromosomal eukaryotic gene or gene fragment (having at least one intron or at least one exon of a eukaryotic gene).

A nucleic acid capable of undergoing homologous recombination with the eukaryotic gene is introduced into the bacterial cell, and the nucleic acid undergoes homologous recombination with the eukaryotic gene or gene fragment. In a particular embodiment, the nucleic acid undergoes homologous recombination with a targeting frequency of at least about 1 in 1,000.

In one embodiment, the expressed recombinase is a double strand break repair recombinase, such as lambda Beta or other single-stranded DNA binding protein; lambda Exo, or lambda Gam. In another embodiment, the extrachromosomal eukaryotic gene or gene fragment may be located on a bacterial artificial chromosome, yeast artificial chromosome, P1 artificial chromosome, plasmid or cosmid. In yet another embodiment, the eukaryotic gene or gene fragment is derived from a mammalian organism, such as a mouse or human.

In several additional embodiments, the nucleic acid undergoing homologous recombination may encode a recombinase, functional fragments or variants of a recombinase, or an epitope tag.

Also disclosed are methods of altering intrachromosomal DNA of a eukaryotic cell. In these methods, an altered eukaryotic gene or gene fragment is introduced into the eukaryotic cell. The introduced eukaryotic gene or gene fragment has been altered by homologous recombination using the methods of this disclosure.

For example, extrachromosomal DNA including the eukaryotic gene or gene fragment is introduced into a bacterial cell having an intrachromosomal gene encoding a recombinase operably linked to a de-repressible promoter. The bacterial cell is then induced to express the recombinase. A nucleic acid molecule capable of undergoing homologous recombination with the eukaryotic gene or gene fragment is introduced into the bacterial cell. The eukaryotic gene or gene fragment undergoes homologous recombination with the nucleic acid, and altered eukaryotic gene or gene fragment may then be isolated and introduced into a eukaryotic cell.

In one embodiment, the eukaryotic gene or gene fragment introduced into the eukaryotic cell is located on a bacterial artificial chromosome. The eukaryotic gene or gene fragment is capable of undergoing homologous recombination with a target gene in the cell, thereby altering the nucleic acid sequence of the eukaryotic cell's intrachromosomal DNA. In specific, non-limiting examples, the eukaryotic cell is a mammalian cell, an embryonic stem cell, or a zygote.

Also disclosed are mutant mammals which have had one or more of their genes altered by homologous recombination with a bacterial artificial chromosome carrying a eukaryotic gene or gene fragment that has been altered by the disclosed methods. The gene alteration can introduce a recombinase into the mutant mammal, such as a site-specific recombinase.

A mobilizable lambda DNA is also disclosed herein that is isolated as a mini-lambda prophage. The mobilizable lambda DNA can be transformed into any bacterial strain of interest. The lambda DNA integrates into the bacterial chromosome to generate a defective prophage that expresses the recombinase.

The present disclosures provide methods of enhancing the efficiency of homologous recombination. The disclosures will be better understood by reference to the following explanation of terms used and detailed description of methods for carrying out the invention.

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

**Explanation of terms used**

**3' overhang:** two nucleic acid sequences which when annealed are partially double-stranded and partially single-stranded. The single-stranded end or ends extend away from the double-stranded segment in a 5' to 3' direction.

5       **5' overhang:** two nucleic acid sequences which when annealed are partially double-stranded and partially single-stranded. The single-stranded end or ends extend away from the double-stranded segment in a 3' to 5' direction.

**Arabinose:** A simple 5-carbon sugar metabolized by *E. coli*. In one embodiment, it is used as a chemical to inactivate repression and to induce and activate expression from the promoter pBAD.

10

**Attachment Site (att):** A site specific site for recombination that occurs on either a phage or a chromosome. An attachment site on lambda is termed "attP", while an attachment site of a bacterial chromosome is "attB." Integrase mediated recombination of an attP site with an attB site leads to integration of the  $\lambda$  prophage in the bacterial chromosome.

15

**Bacterial artificial chromosome (BAC):** bacterial artificial chromosomes (BACs) have been constructed to allow the cloning of large DNA fragments in *E. coli*, as described in O'Conner et al., Science 244: 1307-12, 1989; Shizuya et al., Proc. Natl. Acad. Sci. USA 89: 8794-7, 1992; Hosoda et al., Nucleic Acids Res. 18: 3863-9, 1990; and Ausubel et al. This system is capable of stably propagating mammalian DNA over 300 kb. In one embodiment, a BAC carries the F replication and partitioning systems that ensure low copy number and faithful segregation of plasmid DNA to daughter cells. Large genomic fragments can be cloned into F-type plasmids, making them of use in constructing genomic libraries.

20

**Beta:** the 28 kDa lambda Beta ssDNA binding polypeptide (and nucleic acid encoding lambda beta) involved in double strand break repair homologous recombination. DNA encoding Beta (*bet*) and polypeptide chains having lambda Beta activity are also referred to herein as *bet*. See Examples 1 and 14 and references therein

25

for discussion. The lambda Beta protein binds to single-stranded DNA and promotes renaturation of complementary single strand regions of DNA. See Karakousis et al, J. Mol. Biol. 276: 721-733, 1998; Li et al., J. Mol. Biol. 276: 721-733, 1998; Passy et al., PNAS 96: 4279-4284, 1999.

5           Functional fragments and variants of Beta include those variants that maintain their ability to bind to ssDNA and mediate the recombination function of lambda Beta as described herein, and in the publications referenced herein. It is recognized that the gene encoding Beta may be considerably mutated without materially altering the ssDNA binding function or homologous recombination function of lambda Beta. First, the genetic code is well-known to be degenerate, and thus different codons encode the same amino acids. Second, even where an amino acid mutation is introduced, the mutation may be conservative and have no material impact on the essential functions of lambda Beta. See Stryer, Biochemistry 3rd Ed., (c) 1988. Third, part of the lambda Beta polypeptide chain may be deleted without impairing or eliminating its ssDNA binding protein function, or its recombination function. Fourth, insertions or additions may be made in the lambda Beta polypeptide chain -- for example, adding epitope tags -  
10           - without impairing or eliminating its essential functions. Ausubel et al., 1997.

**Biolistics:** Insertion of DNA into cells using DNA-coated micro-projectiles. Also known as particle bombardment or microparticle bombardment. The approach is further described and defined in U.S. Pat. No. 4,945,050, which is herein incorporated by reference.  
20           

**cDNA (complementary DNA):** A piece of DNA lacking internal, non-coding segments (introns) and regulatory sequences that determine transcription. cDNA may be synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.  
25           

**Cosmid:** Artificially constructed cloning vector containing the *cos* gene of phage lambda. Cosmids can be packaged in lambda phage particles for infection into *E. coli*; this permits cloning of larger DNA fragments (up to 45 kb) than can be introduced into bacterial hosts in *plasmid* vectors.



**Cre:** The Cre recombinase is a site-specific recombinase. It recognizes and binds to specific sites called *loxP*. Two *loxP* sites recombine at nearly 100% efficiency in the presence of Cre, thus, permitting DNA cloned between two such sites to be efficiently removed by the Cre-mediated recombination.

5       **De-repressible Promoter:** When a repressor is bound to a de-repressible promoter transcription is substantially decreased as compared to transcription from the de-repressible promoter in the absence of repressor. By regulating the binding of the repressor, such as by changing the environment, the repressor is released from the de-repressible promoter, and transcription increases. In this definition, a de-repressible  
10 promoter does not require an activator for transcription. One specific, non-limiting example is the PL promoter, which is regulated by the repressor *cI*, but is not activated by an activator. (The arabinose promoter is not a simple de-repressible promoter as arabinose inactivates the repressor *AraC* and converts it to an activator).

In one embodiment, the de-repressible promoter is a temperature sensitive de-repressible promoter. For example, by increasing the temperature, the repressor is  
15 released from the promoter, or can no longer bind to the promoter with a high affinity, and transcription is increased from the promoter. One specific, non-limiting example is the induction of PL promoter activity by increasing the temperature of the cell. Increased temperature inactivates the temperature-sensitive repressor *cI*, allowing genes  
20 that are operably linked to the PL promoter to be expressed at increased levels. One of skill in the art can readily identify a repressible promoter.

In one embodiment, a de-repressible promoter is auto-regulated. One specific, non-limiting example of an auto-regulated de-repressible promoter is PL. If only one copy of a gene encoding *cI* is present, yet many copies of the PL promoter are present,  
25 expression of *cI* is upregulated such that transcription is blocked from any of the PL promoters.

**Double-strand break repair recombination:** A type of homologous recombination exemplified by the lambda recombination proteins *Exo*, *Beta* and *Gam*, and shared by numerous other recombinase systems. A double strand break is the

- initiation point for concerted action of recombination proteins. Typically, an exonuclease degrades processively from the 5' ends of these break sites, and ssDNA binding polypeptide binds to the remaining 3' single strand tail, protecting and preparing the recessed DNA for homologous strand invasion (Szostak et al., Cell 33: 25-35, 1983; Little, J. Biol. Chem. 242: 679-686, 1967; Carter et al., J. Biol. Chem. 246: 2502-2512, 1971; Lindahl et al., Science 286: 1897-1905, 1999). Examples of ssDNA binding polypeptides which bind to either ssDNA and/or dsDNA with 3' overhangs and promote double-strand break repair recombination include lambda Beta, RecT of *E. coli*, Erf of phage p22, and Rad52 in various eukaryotic cells including yeast and mammalian cells.

**Electrocompetent:** cells capable of macromolecular uptake upon treatment with electroporation.

- Electroporation:** a method of inducing or allowing a cell to take up macromolecules by applying electric fields to reversibly permeabilize the cell walls. Various methods and apparatuses used are further defined and described in: U.S. Pat. No. 4,695,547; U.S. Pat. No. 4,764,473; U.S. Pat. No. 4,882,28; and U.S. Pat. Nos. 4,946,793; 4,906,576; 4,923,814; and 4,849,089, all of which are herein incorporated by reference.

- Eukaryotic cell:** a cell having an organized nucleus bounded by a nuclear membrane. These include lower organisms such as yeasts, slime molds, and the like, as well as cells from multicellular organisms such as invertebrates, vertebrates, and mammals. They include a variety of tissue types, such as: endothelial cell, smooth muscle cell, epithelial cell, hepatocyte, cells of neural crest origin, tumor cell, hematopoietic cell, immunologic cell, T cell, B cell, monocyte, macrophage, dendritic cell, fibroblast, keratinocyte, neuronal cell, glial cell, adipocyte, myoblast, myocyte, chondroblast, chondrocyte, osteoblast, osteocyte, osteoclast, secretory cell, endocrine cell, oocyte, and spermatocyte. These cell types are described in standard histology texts, such as McCormack, Introduction to Histology, (c) 1984 by J.P. Lippincott Co.; Wheater et al., eds., Functional Histology, 2nd Ed., (c) 1987 by Churchill Livingstone;

Fawcett et al., eds., Bloom and Fawcett: A Textbook of Histology, (c) 1984 by William and Wilkins, all of which are incorporated by reference in their entirety.

**Exo:** the exonuclease of lambda (and the nucleic acid encoding the exonuclease protein) involved in double strand break repair homologous recombination. See  
5 Example 1 and references therein for discussion.

**Gam:** a lambda protein (and nucleic acid encoding Gam) involved in double strand break repair homologous recombination. It is believed to inhibit cellular nuclease activity such as that encoded by the recBCD and sbcC system of *E. coli*. See Examples 1, 7 and 14 and references therein for discussion. Gam function, when expressed in the  
10 cell, is extremely toxic to the cell, and prevents growth. For this reason tight controls over its expression are always required. As described herein, PL and cI 857 are able to regulate Gam expression

Functional fragments and variants of Exo and Gam: As discussed for Beta (see "Functional fragments And Variants Of Beta"), it is recognized that genes encoding  
15 Exo or Gam may be considerably mutated without materially altering their function, because of genetic code degeneracy, conservative amino acid substitutions, noncritical deletions or insertions, etc. Unless the context makes otherwise clear, the term lambda Exo, *Exo*, or lambda exonuclease are all intended to include the native lambda exonuclease, and all fragments and variants of lambda exonuclease.

**Extrachromosomal:** not incorporated into the chromosome or chromosomes of  
20 a cell. In the context of nucleic acids, extrachromosomal indicates an DNA oligonucleotide that is not covalently incorporated into the chromosome or chromosomes of a cell. **Intrachromosomal** refers to material such as an oligonucleotide that is incorporated into the chromosome or chromosomes of a cell,  
25 such as a DNA oligonucleotide covalently incorporated into the chromosomal DNA of a cell.

**Homologous arm:** nucleotides at or near 5' or 3' end of a polynucleotide which are identical or similar in sequence to the target nucleic acid in a cell, and capable of

mediating homologous recombination with the target nucleic acid. Homologous arms are also referred to as homology arms. In one embodiment, a homology arm includes at least 20 bases of a sequence homologous to a nucleic acid of interest. In another embodiment, the homology arm includes at least 30 base pairs of a sequence homologous to a nucleic acid of interest. In yet another embodiment, a homology arm includes at least 40 base pairs of a sequence homologous to a nucleic acid of interest. In a further embodiment, a homology arm includes from about 50 to about 100 base pairs of a sequence homologous to a nucleic acid of interest.

**Homologous recombination:** an exchange of homologous polynucleotide segments anywhere along a length of two nucleic acid molecules.

**Host cell:** a cell that is used in lab techniques such as DNA cloning to receive exogenous nucleic acid molecules. In one embodiment a host cell is used to maintain or allow the reproduction of a vector, or to facilitate the manipulation of nucleic acid molecules *in vitro*. A host cell can be a prokaryotic or a eukaryotic cell.

**HVJ-mediated gene transfer:** a method of macromolecular transfer into cells using inactivated hemagglutinating virus of Japan and liposomes, as described in Morishita et al., J. Clin. Invest. 91: 2580-2585, 1993; Morishita et al., J. Clin. Invest. 94: 978-984, 1994; which are herein incorporated by reference.

**Inducible promoter:** a promoter whose activity may be increased (or that may be de-repressed) by some change in the environment of the cell. Examples of inducible promoters abound in nature, and a broad range of environmental or hormonal changes may activate or repress them.

**Isolated:** An "isolated" biological component (such as a nucleic acid or protein) has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, *i.e.*, other chromosomal and extra-chromosomal DNA and RNA, and proteins. Thus, nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins

prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

**Linear plasmid vector:** a DNA sequence (1) containing a bacterial plasmid origin of replication, (2) having a free 5' and 3' end, and (3) capable of circularizing and replicating as a bacterial plasmid by joining its free 5' and 3' ends. Examples of  
5 linear plasmid vectors include the linearized pBluescript vector and linearized pBR322 vectors described herein.

**Lipofection:** The process of macromolecular transfer into cells using liposomes. See U.S. Patent No. 5,651,981, which is herein incorporated by reference.

10 **Mini lambda:** A derivative of lambda ( $\lambda$ ) wherein most of the viral lytic genes, including those required for replication and lysis, are deleted. A mini-lambda maintains the red functions (Beta, Exo, and Gam) for homologous recombination and maintains the integration/excision functions (e.g. att, integrase (int). and excisionase (xis)) to insert and excise its DNA from the chromosome.

15 **Nucleic acid:** a deoxyribonucleotide or ribonucleotide polymer in either single or double stranded form, including known analogs of natural nucleotides unless otherwise indicated.

**Oligonucleotide:** a single-stranded nucleic acid ranging in length from 2 to about 500 bases, for example, polynucleotides that contain at least 20 or 40 nucleotides.  
20 Oligonucleotides are often synthetic but can also be produced from naturally occurring polynucleotides.

**Operably linked:** A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter  
25 is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

**Phagemid artificial chromosome:** Also referred to as P1 artificial chromosome. A type of artificial chromosome allowing for stable cloning of very large DNA fragments. Further described in Shepherd, et al., Proc. Natl. Acad. Sci. USA, 92: 2629, 1994; Iannou et al., Nature Genetics, 6: 84-89, 1994.

5       **Phage-based recombination systems:** Bacteria such as *E. coli* encode their own homologous recombination systems, which are used in repair of DNA damage and to maintain a functional chromosome. The viruses or phages that inhabit bacteria often carry their own recombination functions. Phage  $\lambda$  carries the Red recombination system. These phage systems can work with the bacterial recombination functions or  
10       independently of them.

**PL promoter:** the major leftward promoter of bacteriophage lambda. Once the lambda DNA is incorporated into the bacterial chromosome, transcription from this promoter is substantially repressed by the cI repressor. Upon inactivation of the cI repressor, for example by heat shock of a temperature sensitive mutant, transcription  
15       from the PL promoter is activated, leading to expression of lambda genes. See Fig. 1; Sambrook et al. Ch. 2, Stryer, and Court and Oppenheim.

**Plasmid:** Autonomously replicating, extrachromosomal DNA molecules, distinct from the normal bacterial genome and nonessential for bacterial cell survival under nonselective conditions.

20       **Polynucleotide:** A double stranded or single stranded nucleic acid sequence of any length. Therefore, a polynucleotide includes molecules which are 15, 50, 100, 200 nucleotides long (oligonucleotides) and also nucleotides as long as a full length cDNA.

**Polypeptide:** any chain of amino acids, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation).

25       **Prokaryote:** Cell or organism lacking a membrane-bound, structurally discrete nucleus and other subcellular compartments.

**Probes and primers:** A nucleic acid probe comprises an isolated nucleic acid attached to a detectable label or reporter molecule. Typical labels include radioactive

isotopes, ligands, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, *e.g.*, in Sambrook *et al.*, (1989) and Ausubel *et al.*, (1997).

Primers are short nucleic acids, preferably DNA oligonucleotides 15 nucleotides or more in length. Primers may be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand. The 3' hydroxyl end of the primer may be then extended along the target DNA strand through the use of a DNA polymerase enzyme. Primer pairs (one on either side of the target nucleic acid sequence) can be used for amplification of a nucleic acid sequence, *e.g.*, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in Sambrook *et al.*, (1989), Ausubel *et al.*, (1987). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA). Under appropriate conditions, the specificity of a particular probe or primer increases with its length. Thus, in order to obtain greater specificity, probes and primers may be selected that comprise 20, 25, 30, 35, 40, 50 or more consecutive nucleotides of related cDNA or gene sequence.

**Purified:** The term purified does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified lambda Beta preparation or ssDNA binding polypeptide is one in which the protein is more enriched than the protein is in its natural environment within a cell. Preferably, a preparation of lambda Beta is purified such that the polypeptide represents at least 50% of the total protein content of the preparation.

**RecA:** The RecA protein is a central protein that has an activity as in the recombination function of *E. coli*. Homologues are found in all other organisms. RecA protein allows two homologous DNAs to find each other among non-homologous DNAs and then trade or transfer strands with each other. This exchange occurs by

RecA binding to a single stranded region in one of the DNAs and using that strand to search for its dsDNA homolog, binding to the dsDNA and causing the single strand to pair with its complement in the dsDNA ultimately displacing the identical strand of the duplex. This strand transfer generates a key intermediate in the RecA-mediated

5 recombination process.

***recE recT* genes and the Rac prophage:** *E. coli* and other bacteria contain in their chromosomes remnants of viruses. These viruses or prophages are for the most part defective and may contain only a few genes of the original virus. In *E. coli*, one defective prophage is called Rac. Two genes, *recE* and *recT* of the Rac prophage,  
10 encode homologous recombination functions. These genes are normally silent but the *sbcA* mutation activates their constitutive expression. Thus, the *sbcA* mutant is active for recombination.

**Recombinases:** proteins that, when included with an exogenous targeting polynucleotide, provide a measurable increase in the recombination frequency between  
15 two or more oligonucleotides that are at least partially homologous.

**Selection markers or selectable markers:** nucleic acid sequences which upon intracellular expression are capable of conferring either a positive or negative selection marker or phenotypic characteristic for the cell expressing the sequence. The term "selection marker" or "selectable marker" includes both positive and negative selection  
20 markers. A "positive selection marker" is a nucleic acid sequence that allows the survival of cells containing the positive selection marker under growth conditions that kill or prevent growth of cells lacking the marker. An example of a positive selection marker is a nucleic acid sequence which promotes expression of the neomycin resistance gene. Cells not containing the neomycin resistance gene are selected against  
25 by application of G418, whereas cells expressing the neomycin resistance gene are not harmed by G418 (positive selection). A "negative selection marker" is a nucleic acid sequence that kills, prevents growth of or otherwise selects against cells containing the negative selection marker, usually upon application of an appropriate exogenous agent. An example of a negative selection marker is a nucleic acid sequence which promotes



expression of the thymidine kinase gene of herpes simplex virus (HSV-TK). Cells expressing HSV-TK are selected against by application of ganciclovir (negative selection), whereas cells not expressing the gene are relatively unharmed by ganciclovir. The terms are further defined, and methods further explained, by U.S. Patent 5,464,764, which is herein incorporated by reference.

**Selectable phenotype:** A cell with a selectable phenotype is one that expresses a positive or negative selection marker.

**Sequence identity:** the similarity between two nucleic acid sequences, or two amino acid sequences is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar are the two sequences.

Methods of alignment of sequences for comparison are well-known in the art. Various programs and alignment algorithms are described in: Smith and Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman and Wunsch, *J. Mol. Bio.* 48:443, 1970; Pearson and Lipman, *Methods in Molec. Biology* 24: 307-331, 1988; Higgins and Sharp, *Gene* 73:237-244, 1988; Higgins and Sharp, *CABIOS* 5:151-153, 1989; Corpet et al., *Nucleic Acids Research* 16:10881-90, 1988; Huang et al., *Computer Applications in BioSciences* 8:155-65, 1992; and Pearson et al., *Methods in Molecular Biology* 24:307-31, 1994. Altschul et al. (1994) presents a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul et al., *J. Mol. Biol.* 215:403-410, 1990) is available from several sources, including the National Center for Biological Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. It can be accessed at the NCBI website, together with a description of how to determine sequence identity using this program.

Homologues of lambda Beta, Exo and Gam, and ssDNA binding proteins typically possess at least 60% sequence identity counted over full-length alignment with the amino acid sequence of the protein being evaluated (that is, lambda Beta, Exo or Gam, or ssDNA binding protein such as P22 Erf, RecT, and Rad52) using the NCBI Blast 2.0, gapped blastp set to default parameters. For comparisons of amino acid sequences of greater than about 30 amino acids, the Blast 2 sequences function is employed using the default BLOSUM62 matrix set to default parameters, (gap existence cost of 11, and a per residue gap cost of 1). When aligning short peptides (fewer than around 30 amino acids), the alignment should be performed using the Blast 2 sequences function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties). Proteins with even greater similarity to the reference sequence will show increasing percentage identities when assessed by this method, such as at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 98%, or at least 99% sequence identity. When less than the entire sequence is being compared for sequence identity, homologs will typically possess at least 75% sequence identity over short windows of 10-20 amino acids, and may possess sequence identities of at least 85% or at least 90% or 95% depending on their similarity to the reference sequence. Methods for determining sequence identity over such short windows are described at the NCBI website

One of skill in the art will appreciate that these sequence identity ranges are provided for guidance only; it is entirely possible that strongly significant homologs or other variants could be obtained that fall outside of the ranges provided.

**Single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA):**

ssDNA is DNA in a single polynucleotide chain; the DNA bases are not involved in Watson-Crick base pairing with another polynucleotide chain. dsDNA involves two or more complementary polynucleotide chains, in which the two polynucleotide chains are at least partially Watson-Crick base-paired to each other. Note that a segment of DNA may be partially ssDNA and partially dsDNA, for example if there are gaps in one polynucleotide chain of a segment of dsDNA, or there are 5' or 3' overhangs. ssDNA and dsDNA may contain nucleotide analogs, nonnaturally occurring or synthetic

nucleotides, biotin, or epitope or fluorescent tags. ssDNA or dsDNA may be labeled; typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

**Site-specific recombinase:** a recombinase whose activity is limited to DNA of a specific sequence. Examples include the Cre, FLP and FLPE recombinases. Lambda Int is specific for the att sites. Cre recombinases are site-specific for *loxP* recombination sites in a DNA sequence, whereas FLP and FLPE recombinases are site-specific for *FRT* recombination sites. A recombination site is a nucleic acid sequence specifically recognized by a recombinase. For example, the Cre recombinase specifically binds a *loxP* recombination site, and thereby induces recombination.

**Target nucleic acid sequence:** the nucleic acid segment which is targeted for homologous recombination. Typically, this is a segment of chromosomal or extrachromosomal DNA in a cell. Extrachromosomal DNA harboring target nucleic acid sequences may include episomal DNA, plasmid DNA, bacterial artificial chromosome, phagemid artificial chromosomes, yeast artificial chromosomes, cosmids, and the like. The target nucleic acid sequence usually harbors a gene or gene fragment which will be mutated in some fashion upon homologous recombination. Examples of target nucleic acid sequences include DNA sequences surrounding the tyr 145 UAG amber mutation of galK, as described in Yu et al., PNAS 97: 5798-5983, 2000, and in Example 3 of this application; the second exon of mouse hox 1.1 gene, as described in U.S. Patent No. 5,464,764; the human hemoglobin S gene mutation as described in Example 15 of this application.

**Targeting frequency:** The frequency with which a target nucleic acid sequence undergoes homologous recombination. For example, extrachromosomal DNA is introduced into a eukaryotic cell. The extrachromosomal DNA has sequences capable of undergoing homologous recombination with a target intrachromosomal DNA sequence. After introducing the extrachromosomal DNA and allowing homologous recombination to proceed, the total number of cells may be determined, and the number of cells having the target DNA sequence altered by homologous recombination may be

determined. The targeting frequency is the number of cells having the target DNA sequence altered, divided by the total number of cells. For example, if there are a total number of one million cells, and 1,000 of these cells have the target DNA sequence altered, then the targeting frequency is 1 in 1,000, or  $10^{-3}$ .

5       **Transformed:** As used herein, the term transformation encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of DNA (including DNA linked to Beta protein) by electroporation, lipofection, and biolistics.

10       **Upstream:** refers to nucleic acid sequences that precede the codons that are transcribed into a RNA of interest. Similarly, downstream refers to nucleic acid sequences that follow codons that are transcribed into a RNA of interest.

15       **Variants of Amino Acid and Nucleic Acid Sequences:** The production of lambda Beta, Exo or Gam, or other ssDNA binding polypeptide can be accomplished in a variety of ways. DNA sequences which encode for the protein, or a fragment of the protein, can be engineered such that they allow the protein to be expressed in eukaryotic cells, bacteria, insects, and/or plants. In order to accomplish this expression, the DNA sequence can be altered and operably linked to other regulatory sequences. The final product, which contains the regulatory sequences and the nucleic acid encoding the  
20       therapeutic protein, is operably linked into a vector, allowing stable maintenance in a cell. This vector can then be introduced into the eukaryotic cells, bacteria, insect, and/or plant. Once inside the cell, the vector allows the protein to be produced.

25       One of ordinary skill in the art will appreciate that the DNA can be altered in numerous ways without affecting the biological activity of the encoded protein. For example, PCR may be used to produce variations in the DNA sequence which encodes lambda Beta, Exo or Gam, or other ssDNA binding proteins. Such variants may be variants that are optimized for codon preference in a host cell that is to be used to express the protein, or other sequence changes that facilitate expression.

Two types of cDNA sequence variant may be produced. In the first type, the variation in the cDNA sequence is not manifested as a change in the amino acid sequence of the encoded polypeptide. These silent variations are simply a reflection of the degeneracy of the genetic code. In the second type, the cDNA sequence variation  
5 does result in a change in the amino acid sequence of the encoded protein. In such cases, the variant cDNA sequence produces a variant polypeptide sequence. In order to preserve the functional and immunologic identity of the encoded polypeptide, such amino acid substitutions are ideally conservative in highly conserved regions. Conservative substitutions replace one amino acid with another amino acid that is  
10 similar in size, hydrophobicity, etc. Outside of highly conserved regions, non-conservative substitutions can more readily be made without affecting function of the protein. Examples of conservative substitutions are shown in Table 1 below.

Table 1

	Original Residue	Conservative Substitution
	Ala	Ser
	Arg	Lys
5	Asn	Gln, His
	Asp	Glu
	Cys	Ser
	Gln	Asn
	Glu	Asp
10	Gly	Pro
	His	Asn; Gln
	Ile	Leu, Val
	Leu	Ile; Val
	Lys	Arg; Gln; Glu
15	Met	Leu; Ile
	Phe	Met; Leu; Tyr
	Ser	Thr
	Thr	Ser
	Trp	Tyr
20	Tyr	Trp; Phe
	Val	Ile; Leu

Variations in the cDNA sequence that result in amino acid changes, whether conservative or not, should be minimized in order to preserve the functional and immunologic identity of the encoded protein. The immunologic identity of the protein may be assessed by determining whether it is recognized by an antibody to the protein; a variant that is recognized by such an antibody is immunologically conserved. Particular examples of cDNA sequence variants introduce no more than 20, and fewer than 10 amino acid substitutions, into the encoded polypeptide. Variant amino acid

sequences may, for example, be at least 80, 90 or even 95% identical to the native amino acid sequence.

**Yeast artificial chromosome (YAC):** A vector used to clone DNA fragments (up to 400 kb); it is constructed from the telomeric, centromeric, and replication origin sequences needed for replication in yeast cells (see Ausubel et al.).

*Use of the Lambda-encoded Red recombination system  
in recombineering mediated by a defective prophage*

10 Bacteriophage  $\lambda$  contains a homologous recombination system termed Red, which is functionally analogous to the RecET recombination system of *Rac*. Like RecET, Red recombination requires two genes: *red $\alpha$*  or *exo*, which is analogous to *recE*, and *red $\beta$*  (or *bet*), which is analogous to *recT*. Exo is a 5'-3' exonuclease that acts processively on linear dsDNA. Beta binds to the ssDNA overhangs created by Exo and stimulates annealing to a complementary strand but cannot promote direct strand invasion and exchange on its own. The recombination functions of Exo and Beta are again assisted by  $\lambda$  phage-encoded Gam, which inhibits the RecBCD activity of the host cell<sup>47</sup>.  $\lambda$  Red-mediated recombination events are 10 to 1000 times more efficient than those observed in *recBC sbcBC* or *recD* strains. Because homologous recombination is increased dramatically by the addition to the host of phage-encoded protein functions, this procedure is widely applicable to any *E. coli* strain and to other bacterial species as well.

25 A defective  $\lambda$  prophage-based system is disclosed herein for Red-mediated recombineering (see Fig. 2). In this system, Gam, Beta, and Exo are encoded by a defective lambda prophage, which is integrated into the *E. coli* chromosome of a bacterial cell (e.g. *E. coli*) (see Fig. 6 and Fig. 9). Expression of Gam, Beta, and Exo is under the tight control of a de-repressible promoter. In the example shown, the de-repressible PL promoter is under the control of the temperature-sensitive  $\lambda$  *cl857* repressor. At 32°C, when the repressor is active, expression of the PL promoter and

these genes is undetectable. However, when the cells are shifted to 42°C for a brief 15 min period, the repressor is inactivated and the genes are expressed at very high levels. In contrast, promoters that can be activated, which are present on plasmids, are notoriously difficult to control and Red and Gam functions would be expressed even in the absence of the inducer, such as arabinose. Low-level expression of Gam causes a  
5 RecBCD defect, a condition that results in plasmid instability and loss of cell viability.

The tight regulation afforded by the prophage system, combined with the fact that the  $\lambda$  promoter, which drives Gam and Red expression is a very strong promoter, makes it possible to achieve recombination frequencies that are at least 50- fold higher  
10 than those found with the plasmid-based system used previously (see Muyrers et al, *Nucleic Acids Res* 27: 1555-1557, 1999; Yu, D. et al. *Proc Natl Acad Sci U S A* 97: 5978-5983, 2000). and several orders of magnitude higher than previously described strains in which linear recombination has been studied. The prophage itself is genetically stable, unlike plasmids, and does not rely on the presence of drug selection  
15 for maintenance.

FIG. 2 illustrates the design of primers for amplification of a dsDNA recombination cassette, and a strategy for generating recombinant DNA molecules and gene replacement. The steps and advantages of this system are outlined below.

Classical recombinant DNA technology or genetic engineering has primarily  
20 relied upon the presence of restriction enzyme cleavage sites to judiciously cleave DNA and the use of DNA ligase to covalently join different DNAs to make the recombinants wanted. The ability to do genetic engineering has been simplified by the polymerase chain reaction (PCR), which allows restriction sites to be incorporated into linear PCR products thereby allowing more precise positioning of those sites. All genetic  
25 engineering technology breaks down, however, when cloning vehicles and the target contain hundreds of kilobases of DNA. Examples include the bacterial chromosome and large genomic BAC clones. Even rare restriction enzyme sites occur frequently on such large DNA molecules making the effort to use unique sites impossible. Furthermore, the *in vitro* manipulation of linear DNAs of this length is also extremely difficult.



Therefore, once large BAC cloning technology became available in *E. coli*, modification of the BAC clones became the primary problem. Initially a combination of genetic engineering technology and classical homologous recombination techniques were adapted to modify the large genomic clones. Classical homologous recombination in *E. coli* depends upon significant (>500 bp) stretches of homology between DNAs.

Fig. 2A depicts a typical genetic engineering protocol to modify a target on a BAC clone with a cassette and compares that technology with the recently developed recombineering technology that uses special phage recombination functions. In general, there are many steps required for classical engineering, and the final product cannot be engineered as precisely as by the new recombineering technology. The critical advance in the recombineering methodology is the use of phage recombination functions that generate recombination products using homologies of 50 bp (or less). Note that the target homologies in Fig. 2A and Fig. 2B are represented by the striped boxes. In the method outlined in Fig. 2A, those boxes must be at least 500 bp long, whereas in the method outlined in Fig. 2B, they need only be about 40 to about 50 bp long.

Genetic engineering steps to generate BAC recombinant include:

- Cleavage of the cassette DNA by a restriction enzyme.
- Cleavage of target on plasmid by a restriction enzyme.

(Note in this case the vector has been pre-engineered to contain target fragments.)

- Joining the cassette to the plasmid by DNA ligase.
- Transforming the DNA into cells.
- Selection of drug<sup>R</sup> clones and isolation of the plasmid.
- Verification of the cloned cassette.
- Transformation of the cloned cassette into the BAC strain.
- Several recombination steps -> -> -> to introduce the cloned cassette into the BAC.

Recombineering steps to generate BAC recombinants include:

- Generation of two primers (white and black arrows).
- Generation of PCR amplified cassette with flanking homologies (In this panel, exemplary striped homology segments shown are 50 bp long).
- Induction of phage recombination functions into BAC strain or introduction of BAC DNA into strain carrying recombination functions.
- Transformation of cells containing BAC and recombination functions with PCR cassette.
- Generation of recombinant *in vivo*.
- Detection of recombinant:
  - Detection by selection or counterselection.
  - Detection by direct screening (colony hybridization).
  - Detection by detecting a label on the nucleic acid (e.g. when DNA includes a DNA adduct or a marker such as biotin)

As disclosed herein, the defective  $\lambda$  prophage has also been transferred to the BAC host strain DH10B so that it can be used for BAC engineering. The modified DH10B strain called DY380 can be transformed with BAC DNA at efficiencies of  $10^{-6}$  to  $10^{-4}$ . The utility of DY380 cells for BAC engineering has been demonstrated by introducing a 250 kbp mouse BAC that contains the *neuronal-specific enolase 2* (*Eno2*) gene into DY380 cells by electroporation and then modifying the BAC by introducing a Cre-expressing targeting cassette into the 3' end of the *Eno2* gene using Red recombination (see Example 20). The targeting cassette was PCR-amplified from a template plasmid using chimeric 63 nucleotide (nt) primers. The 3' 21 nt of each primer was homologous to the targeting cassette, while the 5' 42 nt was homologous to the last exon of *Eno2* where the cassette was to be targeted (see Fig. 10). DY380 cells were then electroporated with the amplified targeting cassette and correctly targeted colonies were obtained at an efficiency approaching  $10^{-4}$  following the induction of Red expression; no targeted colonies were obtained in uninduced cells.

As also disclosed herein, the modified full length BAC was purified and injected in mouse zygotes and a BAC transgenic line established. Two other transgenic lines carrying a shorter 25 kbp subclone of the modified *Eno2* gene on pBR322 were also

established as controls. The 25 kbp subclone carries the entire modified *Eno2* coding region as well as 10 kbp of 5' flanking sequence and 5 kbp of 3' flanking sequence. The activity of the *cre* gene in the different transgenic lines was then assessed by crossing the mice to ROSA26 reporter mice. These mice carry a lacZ reporter that can be activated by Cre recombinase. In mice carrying the full length BAC transgene, Cre activity was detected in all *Eno2*-positive neurons. In contrast, not all *Eno2*-positive neurons expressed Cre in the transgenic mice carrying the smaller 25 kbp subclone, and the pattern of Cre expression varied between the two different 25 kbp subclone lines. These results are consistent with previous studies showing that regulatory sequences can be located hundreds of kilobases from a gene, and highlight the usefulness of BAC engineering for in this case generating Cre-expressing lines for use in conditional knockout experiments.

Arabinose-inducible *flpe* or *cre* genes have also been introduced into the defective prophage carried in strain DY380. *flpe* is a genetically engineered *flp* that has a higher recombination frequency than the original *flp* (Buchholz et al., *Nat Biotechnol* 16: 657-662, 1998, herein incorporated by reference). The site-specific recombinases Flpe and Cre are important tools used to add or delete DNA segments (e.g. drug cassettes). Flpe and Cre expression can be induced by the addition of arabinose and used to remove the selection marker from the targeted locus. This will be especially important in cases where the selection marker interferes with the expression of the targeted locus. It is important to note that even excision of the selectable marker by Flpe or Cre recombination leaves behind the *frt* or *loxP* site as a scar on the targeted locus.

25        *BAC modification without leaving markers or 'scars' at the target site and direct genomic modification*

A two-step procedure for BAC targeting has been developed wherein many kinds of mutations can be introduced into BACs without leaving behind a selectable marker, such as a drug selection marker, at the targeted locus. In one embodiment, a two-step procedure is utilized. This is exemplified in the following specific, non-

limiting example. a PCR-generated targeting cassette containing a *sacB-neo* fusion gene was targeted to a BAC or other DNA. Cells containing the *sacB-neo* cassette targeted to the genomic DNA of the BAC were then transformed with a second targeting DNA to the same region. This cassette was designed to replace the *sacB-neo* cassette, and in one instance, contained short genomic sequences that carried a more subtle mutation, such as a small insertion. By placing these newly transformed cells on media with 7% sucrose, selective pressure was applied against SacB expression, which converts sucrose to a bacteriotoxin (Muyrers et al., *EMBO Rep* 1: 239-243: 2000). Growth on sucrose plates thus selected for cells that have potentially replaced the *sacB-neo* targeting cassette with the second targeting cassette containing the small insertion. Because spontaneous mutations occur in *sacB* to cause sucrose resistance at frequencies approaching 1 in  $10^4$ , recombinants were identified among sucrose resistant colonies as those that have also become neomycin sensitive. As disclosed herein, by combining the power of Red recombination with selection/countersélection using *sacB-neo*, other kinds of genetic changes besides insertions can also be generated, including deletions and point mutations, and these mutations can be introduced into virtually any large DNA molecule such as a BAC, PAC, or the *E. coli* chromosome without any accompanying selectable marker.

The high frequency of recombination generated by the defective prophage system described herein also makes it possible to modify a bacterial genome or a BAC in a single step without drug selection or countersélection. In one specific, non-limiting example, a 24-bp flag tag was introduced into a 125-kbp BAC directly by recombination, without selection into the 5' end of the *SRY-box containing gene 4* (*Sox4*) (see Example 21). The recombinants were found by screening individual cells from the BAC electroporated culture.

Because homologies involved in Red-mediated recombination can be very short, targeting cassettes can also be made by simply annealing two complementary synthetic ssDNA oligonucleotides together. As described herein, a 70 bp targeting cassette constructed in this manner recombines with the *E. coli* chromosome to create point mutations at frequencies approaching one in a thousand electroporated cells. Point

mutations corresponding to human disease-causing mutations can thus be introduced into any human or mamalian gene carried on a BAC with ease and the affect of this mutation on gene function assayed in a transgenic that carries a null mutation in the corresponding mouse gene.

5

#### *Cloning DNA by gap repair*

Fragments can be subloned from BACs by Red-mediated recombination without the use of restriction enzymes or DNA ligases. Thus, any region of the BAC is amenable to subcloning, and subcloning does not depend on the placement of appropriate restriction enzyme sites. Subcloning relies on gap repair to recombine the free ends of a linear plasmid vector with homologous sequences carried on the BAC. An example is shown in Figure 8 and 10. The linear plasmid vector with, for example, an *amp* selectable marker and an origin of replication carries the recombinogenic ends. The vector is generated, for example, by PCR amplification using two chimeric primers. The 5' end of each primer has homology to the extremities of the BAC sequence to be subcloned; the 3' end of each primer is used to prime and amplify the linear plasmid DNA. Recombination generates a circular plasmid in which the DNA insert is retrieved from the BAC via gap repair. Circular recombinant plasmids are selected by their drug resistance (e.g. Amp<sup>R</sup>) phenotype. Different sizes of fragments that can be subcloned depending on the cloning vector utilized. With a high copy vector such as pBluescript, fragments up to about 25 kbp are subcloned. However, with a lower copy vector such as pBR322 is used, fragments as large as about 80 kbp can be subcloned. These larger fragments were shown to be more accurately expressed in a tissue specific manner (as was the entire BAC clone see above).

25

#### *A mobilizable lambda prophage*

As disclosed herein, recombination functions were expressed from their native location in the *pL* operon of a lambda prophage using the natural  $\lambda$  repressor controls (Fig. 9). However, one limitation of the defective prophage system as disclosed in Figs. 6 and 9 is that BACs under study must be moved into recombination-proficient DY380

30

cells before the BAC can be manipulated. In order to overcome this limitation, a novel prophage derivative has been generated that is isolated as a mini-lambda circle DNA carrying a selectable marker (e.g. a drug-resistance marker such as *tet<sup>R</sup>* cassette) and containing the *exo*, *bet*, and *gam* genes under control of the temperature inducible cI857 repressor (Fig. 12). This mini-lambda can be transformed into any bacterial cell, such as a DH10B cell that carry a BAC. The mini-lambda then integrates at the lambda attachment site to generate the defective prophage. This mobilizable prophage makes it possible to introduce the prophage into BAC-containing DH10B libraries and obviates the need to transfer the BAC to DY380 cells.

10

#### *Recombineering using ssDNA*

Recombineering, or the use of a recombinase to mediate recombination using homology arms sufficient to induce recombination, as disclosed herein, can be performed using single-stranded oligos as the targeting cassette). As described in the Examples below (e.g., see Example 3), in *E. coli*, a single base change has been substituted in the *galK* gene and a 3.3 kbp insertion removed from the *galK* gene using single-stranded oligos. Single-stranded oligos have also been used to cure 5 different *Tn10* insertions at different places on the *E. coli* chromosome. Recombineering using single-stranded oligos is very efficient with up to 6% of the electroporated cells being recombinant. Whereas Exo, Beta, and Gam facilitate recombination of PCR amplified dsDNA cassettes with flanking homologies, only Beta is required for ssDNA recombination (see Fig. 13). Maximum recombination is achieved with oligonucleotides of about 70 bases in length, although oligonucleotides of about forty to sixty bases in length can also be used to achieve recombination, albeit at a 5-fold lower frequency. In one embodiment, ssDNA of about 40 to about 70 nucleotides in length is utilized. In another embodiment, ssDNA of about 70 to about 100 nucleotides in length is utilized. In a further embodiment, a ssDNA of about 70 to about 1,000 nucleotides in length is utilized. Interestingly, Beta-mediated recombination activity is less efficient when ssDNA molecules are about 1,000 bases in length. In yet another embodiment,

25

the ssDNA is labeled, such as with a biotinylated nucleotide, a methylated nucleotide, or a DNA adduct.

Recombination with either of two complementary DNA oligos has revealed that although either strand can be efficiently used for recombination, one strand is more competent for recombination than the other. This strand bias has been examined at several positions around the bacterial chromosome with the result that the preferred strand correlates with the lagging strand of DNA replication for each site tested. Without being bound by theory, these results indicated that strand bias is associated with the replication direction through the region being targeted and that ssDNA recombination occurs efficiently near the replication fork. The process of DNA replication results in transient regions of ssDNA that may be accessible to Beta-mediated annealing of the ssDNA oligo. Although recombination occurs on the leading strand, the increased recombination efficiency of the lagging strand oligos may reflect the increased frequency of single-stranded regions during lagging versus leading strand synthesis (Fig. 13). DNA polymerase and DNA ligase could then complete the joining of the annealed oligo to the lagging strand. Without being bound by theory, the increased frequency of ssDNA recombination probably reflects the fact that ssDNA recombination occurs through a simpler mechanism than dsDNA recombination. The ssDNA recombination may require only annealing of one single-stranded oligo to single-stranded regions in the replicating target DNA. Moreover, ssDNA recombination also occurs in yeast with a strand bias that may also be dependent upon replication. The yeast functions required for this recombination are, however, unknown making the finding that only the Beta function from phage  $\lambda$  is required in *E. coli* that much more significant.

In one specific, non-limiting example, a point mutation was engineered into the mouse *Brca2* carried on a BAC using a 70 nt oligo. The targeting efficiency was several times higher than would be found with dsDNA created by annealing oligos and at least 50 times higher than with dsDNA generated by PCR and containing large regions of nonhomology in their center. A 140 nt oligonucleotide has also been used to introduce a 29 amino acid in-frame deletion into exon11 of the *Brca2* gene and a 1.93

kb deletion into the BAC vector backbone (Swaminathan, S. et al. *Genesis* 29, 14-21, 2001, herein incorporated by reference). Finally, a 164 nt oligo has been used to introduce a 24 bp flag tag into the 5' end of *Brca2*. The targeting efficiency for the 164 nt oligo ( $7.7 \times 10^{-3}$ ) was nearly the same as the targeting efficiency for generating  
5 deletions using 140 nt oligos ( $8.3 \times 10^{-3}$  and  $5.4 \times 10^{-3}$ , respectively).

## EXAMPLES

### EXAMPLE 1

#### Modified Lambda Prophage for Defined Expression of Recombination Proteins

10 The molecular genetics of lambda bacteriophage, including its lytic and lysogenic growth cycles, is described in Sambrook et al., *Bacteriophage Lambda Vectors*, Chapter 2 in *Molecular Cloning: a Laboratory Manual*, 2nd Ed., (c) 1989 (hereinafter Sambrook et al., Ch. 2); Stryer, *Control of Gene Expression in Prokaryotes*, Chapter 32 in *Biochemistry* 3rd Ed., pp. 799-823, (c) 1988 (hereinafter Stryer); and  
15 Court and Oppenheim, pp. 251-277 in Hendrix et al. eds., *Lambda II*, Cold Spring Harbor Lab Press, (c) 1983 (hereinafter Court and Oppenheim). The complete sequence of lambda is known (see Genbank Accession No. NC 001416, herein incorporated by reference.)

Phage lambda has a well-characterized homologous recombination system.  
20 Double strand breaks in DNA are the initiation sites for this recombination (Thaler et al., *J. Mol. Biol.* 195: 75-87, 1987). Lambda exonuclease (Exo) degrades processively from the 5' ends of these break sites, and lambda Beta binds to the remaining 3' single strand tail, protecting and preparing the recessed DNA for homologous strand invasion (Little, *J. Biol. Chem.* 242: 679-686, 1967; Carter et al., *J. Biol. Chem.* 246: 2502-2512,  
25 1971).

The lambda recombination system containing *exo* and *bet* without *gam* is efficient at gene replacement using linear substrates with homology arms of more than 1,000 bp in a strain lacking RecBCD nuclease (Murphy, *Journal of Bacteriology* 180:



2063-2071, 1998). To test homology arms of less than 100 bp long as substrates for lambda-mediated recombination, a lambda prophage was modified to express high levels of phage recombination functions for a defined amount of time.

FIG. 1 depicts the defective  $\lambda$  prophage on the *E. coli* chromosome. The defective prophage contains  $\lambda$  genes from *cI* to *int*. The PL operon is intact and expressed under control of the temperature-sensitive lambda cI-repressor (allele *cI857*). A deletion (dotted line) removes the right side of the prophage from *cro* through *attR* and including *bioA* (Patterson et al., Gene 132: 83-87, 1993). On the chromosome, the *nadA* and *gal* operons are to the left of the prophage, and the *bio* genes without *bioA* are to the right. Genes of the  $\lambda$  prophage are shown on the solid line, and genes of the host are shown on the broken line. PL and PR indicate the early left and right promoters of  $\lambda$ . *attL* and *attR* indicate the left and right attachment sites of  $\lambda$ . The lambda genes and functions are described in Sambrook et al., chapter 2, Stryer, and Court and Oppenheim.

The absence of *cro*-repressor allows PL operon expression to be fully derepressed when the temperature sensitive cI-repressor is inactivated at 42°C. The *cro* to *bioA* deletion removes the replication and lytic genes of the prophage. The functions encoded by these lytic genes are toxic to the cell and cause cell death within 7 min after a normal prophage induction. Functions present in the PL operon are also toxic but kill cells only after 60 min of continuous induction (Greer, Virology 66: 589-604, 1975; Kourilsky et al., Biochimie 56: 1517-1523, 1974). Thus, shifting of cells containing the PL operon construct from repressed conditions at 32°C to induced conditions at 42°C allows PL operon expression. Shifting the cells back from 42°C to 32°C (or lower) within 60 min reestablishes repression and prevents cell death.

As the following examples will demonstrate, this modified lambda prophage has produced unexpected advantages in mediating homologous recombination. These unexpected advantages include surprisingly high recombination efficiency, precise control of recombination functions, effective recombination with short homology arms, and ability to generate homologous recombinants with polynucleotides other than long double-stranded DNA. Without wishing to be bound by a single explanation of the

observed effects, it is likely that these unexpected advantages accrue from incorporation of the lambda red genes on the prophage in their native context, thereby limiting the number of copies of the lambda recombination genes. In addition, use of the PL promoter confers the ability to precisely control the timing and production of large  
5 amounts of lambda recombination gene expression.

This system is not limited to expression in *E. coli*, but can also work in other bacteria, such as *Salmonella* and others. It can also work in eukaryotic cells, such as yeast or mammalian cells, with selection of appropriate promoters, and with other modifications of the present to allow expression of the lambda recombinase genes. In  
10 one specific, non-limiting example, in another bacteria, genes between *gam* and *N* (including *N* but not *gam*) can be deleted to remove transcription terminators.

Although the PL promoter is used to illustrate the invention, other de-repressible, inducible or constitutive promoters could be used. Specific, non-limiting examples of inducible promoters are drug inducible promoters (e.g. a tetracycline  
15 inducible promoter) metal inducible promoters (e.g. the metallathione promoter), or a hormone inducible promoter (e.g. a steroid responsive element).

The PL promoter could also be used to drive expression of, for example, P22 genes such as *erf*, or of RecE and RecT.

## EXAMPLE 2

Bacterial strains, Expression of PL operon, Electroporation Methods,  
Identification of Recombinants

5 Bacterial strains used in this work are listed in Table 2.

Table 2

Strains	Genotype
WJW23	<i>his ilv rpsL Δ(argF-lac)U169 nadA::Tn10 gal490 λcl857 Δ(cro-bioA)</i>
ZH1141	<i>W3110 Δ(argF-lac)U169 gal490 λN:lacZ Δ(N-int) cl857 Δ(cro-bioA)</i>
BR3677	<i>lac<sup>+</sup> lacZ(M15) Δ(srl-recA)301::Tn10</i>
DY329	<i>W3110 Δ(argF-lac)U169 nadA::Tn10 gal490 λcl857 Δ(cro-bioA)</i>
DY330	<i>W3110 Δ(argF-lac)U169 gal490 λcl857 Δ(cro-bioA)</i>
DY331	<i>W3110 Δ(argF-lac)U169 Δsrl-recA)301::Tn10 gal490 λcl857 Δ(cro-bioA)</i>
DY378	<i>W3110 λcl857 Δ(cro-bioA)</i>
W3110	"Wild-type"
HME5	<i>Δ(argF-lac)U169 λcl857 Δ(cro-bioA)</i>
HME6	<i>Δ(argF-lac)U169 galK<sub>tyr145UAG</sub> λcl857 Δ(cro-bioA)</i>
HME9	<i>Δ(argF-lac)U169 galK<sub>tyr145UAG</sub> λcl857 Δ(cro-bioA) tyrTV&lt;&gt;cat</i>
HME10	<i>Δ(argF-lac)U169 galK<sub>tyr145UAG</sub> λ cl857 Δ(cro-bioA) tyrTV&lt;&gt;cat Δ(srl-recA)301::Tn10</i>
HME31	<i>Δ(argF-lac)U169 galK&lt;&gt;catsacB λcl857 Δ(cro-bioA)</i>
HME40	<i>Δ(argF-lac)U169 INgal[galM<sup>+</sup>K<sub>tyr145UAG</sub>T<sup>+</sup>E<sup>+</sup>] λcl857 Δ(cro-bioA)</i>
HME43	<i>Δ(argF-lac)U169 galK<sub>tyr145UAG</sub> λ (exo-int)&lt;&gt;cat Δ (gam-N) cl857Δ(cro-bioA)</i>
HME47	<i>galKΠ34&lt;&gt;kan λexo&lt;&gt;cat cl857Δ(cro-bioA)</i>
DY411	<i>galKΠ34&lt;&gt;kan λcl857 Δ (cro-bioA)</i>
DH10B	<i>P mcrA Δ(nrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 deoR recA1 endA1 araD139 Δ(ara, leu)7649 galU galK rspL nupG</i>
DY303	<i>DH10B [λcl857recA*]</i>
DY374	<i>W3110 gal490 nadA::Tn10 [λcl857 Δ(cro-bioA)]</i>
DY363	<i>W3110 ΔlacU169 gal490 [λcl857 (cro-bioA)&lt;tet<sup>r</sup>]</i>
DY380	<i>DH10B [λcl857 (cro-bioA)&lt;tet<sup>r</sup>]</i>
EL11	<i>DH10B [λcl857 (cro-bioA)&lt;cat-sacB]</i>
EL250	<i>DH10B [λcl857 (cro-bioA)&lt;araC-P<sub>BAD</sub>/lpe<sup>b</sup>]</i>
EL350	<i>DH10B [λcl857 (cro-bioA)&lt;araC-P<sub>BAD</sub>cre]</i>

<sup>a</sup> (cro-bioA)<tet indicates substitution of cro-bioA with tet.

<sup>b</sup> P<sub>BAD</sub> represents the promoter of araBAD.

Strain DY329 was constructed by transduction of ZH1141 with P1 phage grown on WJW23, selecting for *nadA::Tn10* tetracycline resistance ( $Tet^R$ ) at 32°C and then screening for the presence of a defective lambda prophage which causes temperature sensitive cell growth at 42 °C. Similar P1 transduction was used to create other strains described in Table 2 using standard media, methods, and selections (Sambrook et al., Molecular Cloning: a Laboratory Manual, 2nd Ed., (c) 1989; Miller, Experiments in Molecular Genetics, Cold Spring Harbor Lab Press (c) 1972). The symbol  $\diamond$  is used to indicate a replacement generated by homologous recombination. The symbol  $\times$  indicates an insertion generated by homologous recombination. A deletion at the point of insertion is indicated in parenthesis following the inserted gene. The entire gal operon in HME40 is inverted (IN).

To induce expression from the PL operon and prepare electroporation-competent cells, overnight cultures grown at 32°C from isolated colonies were diluted 50-fold in LB medium and were grown at 32°C with shaking to an  $OD_{600}$  of about 0.4-0.8. Induction was performed on a 10 ml culture in a baffled conical flask (50 ml) by placing the flask in a water bath at 42°C with shaking (200 revolutions/min) for 15 min. Immediately after the 15 min induction, the flask was swirled in an ice water slurry to cool for 10 min. An uninduced control culture, maintained at 32°C throughout, was also placed into the ice slurry. The cooled 10 ml cultures were centrifuged for 8 min at 5,500 x g at 4°C. Each cell pellet was suspended in 1 ml of ice-cold sterile water, transferred to a 1.5 ml plastic microcentrifuge tube, and was spun for 20 seconds at 4°C at maximum speed in a microcentrifuge. After washing the cell pellets as described two more times, the cells were suspended in 100  $\mu$ l of ice cold sterile water. This volume of competent cells is sufficient for two standard electroporation reactions ( $\sim 10^8$  cells per reaction). Larger cultures can be prepared for a greater number of reactions or for storage of electrocompetent cells at -80°C with 12% glycerol present. Fresh competent cells give highest efficiencies of recombination. To transform cells by electroporation, purified linear donor DNA (1 to 10  $\mu$ l) was mixed

with competent cells in a final volume of 50  $\mu$ l on ice, and then pipetted into a pre-cooled electroporation cuvette (0.1 cm). The amount of donor DNA used per reaction (usually 1 to 100 ng) is indicated for relevant experiments. Electroporation was performed using a Bio-Rad Gene Pulser set at 1.8 kV, 25  $\mu$ F with Pulse controller of 200 ohms. Two protocols have been used interchangeably to allow segregation of recombinant from parental chromosomes within the electroporated cells. In both protocols, the electroporated cells were immediately diluted with 1 ml of LB medium. In one, the cells were incubated for 1 to 2 hours at 32°C before selecting for recombinants. In the other, the cells were immediately diluted and spread on sterile nitrocellulose filters (100 mm) on LB agar. After a 2 hour incubation at 32°C, the filters were transferred to the appropriate agar plates required to select for recombinants. Aliquots were also directly spread on LB agar and incubated at 32°C to determine and examine total viable cells after electroporation. For drug resistant selection, each ml of LB medium contained 10  $\mu$ g of chloramphenicol, 12.5  $\mu$ g of tetracycline, 20  $\mu$ g of kanamycin, 30  $\mu$ g of ampicillin, or 50g of spectinomycin.

Although recombinants were verified by more than one method, the primary detection was for an altered phenotype caused by the modified target gene. Disruption or mutation of the *galK* gene was confirmed by the presence of white colonies on MacConkey galactose indicator agar, disruption of the *rnc* gene for the endoribonuclease RNaseIII was confirmed by the inability of lambdoid type phage to lysogenize (Court, pp. 71-116 in Belasco et al., eds., Control of Messenger RNA Stability, (c) 1993, Academic Press, New York), and deletion of *gam*, *kil*, and *cIII* in the PL operon was scored as an ability of the  $\lambda$  lysogen to survive growth at 42°C (Court and Oppenheim; Greer, Virology 66: 589-604, 1975). PCR analysis was used to confirm the altered structure caused by replacement of a gene. Southern hybridization analyses of parental and recombinant DNAs confirmed structural changes, and DNA from the recombinant clones can be amplified by PCR and sequenced.

In addition to electroporation, any suitable method for macromolecular transfer into bacterial cells would be effective for practicing the methods herein disclosed. For

example, such methods may include exposure to divalent cations, DMSO and the like as described in a variety of standard laboratory publications, such as Sambrook et al. (see particularly pages 1.74-1.84) and Ausubel et al., eds., Current Protocols In Molecular Biology, John Wiley & Sons (c) 1998 (hereinafter Ausubel et al.), herein incorporated  
5 in their entirety.

### EXAMPLE 3

#### Homologous recombination with short linear DNA fragments

The recombination system described in Examples 1 and 2 were used to generate  
10 a single bp mutation in the bacterial *galK* gene.

The *galK* gene encodes a galactokinase that phosphorylates galactose and its derivatives. The *galK* galactokinase phosphorylates 2-deoxygalactose to generate 2-deoxygalactose phosphate (2DGP). While unphosphorylated 2-deoxygalactose has no impact on cell growth, 2DGP is a nonmetabolized sugar phosphate that inhibits cell  
15 growth. Thus, cells containing a wild-type *galK* gene grow poorly on 2-deoxygalactose, a phenotype referred to herein as Gal<sup>+</sup>. In contrast, mutants defective in *galK* grow well in the presence of 2-deoxygalactose (Dog), and have a phenotype referred to herein as DogR - (Adhya, pages 1503-1512 in Escherechia coli and Salmonella typhimurium: Cellular and Molecular Biology, Neihardt et al. eds.,  
20 American Society of Microbiology, 1987). The DogR-phenotype enables ready selection of cells harboring successful recombination events.

Complementary 70 base pair oligonucleotides were synthesized and annealed to each other. The annealed DNA was homologous to an internal coding segment of the bacterial *galK* gene, except that a UAU codon (TYR-145) was changed to a UAG  
25 amber codon. A homologous recombination event between this 70 base pair DNA fragment and the *galK* gene introduces a premature stop codon in the *galK* gene, and is referred to herein as the *galK* amber mutation. This mutation produces a truncated *galK* gene product that lacks function.

The 70-bp DNA fragment was transferred by electroporation into *galK*<sup>+</sup> cells

(HME5) that had been induced for lambda PL operon expression by growth at 42°C for 15 min. After electroporation with the mutant DNA (100 ng) or mock electroporation without DNA, the cells were spread on minimal 0.4% glycerol agar medium with 0.2% 2-deoxygalactose present. Spontaneous resistant mutants occurred frequently ( $10^{-4}$ ) in the absence of mutant DNA. Despite this, the addition of the mutant DNA enhanced the frequency of resistant mutants dramatically, generating one mutant per 500 electroporated cells.

To determine that temperature induction was required, another batch of cells that had not been induced for recombination function was tested in the same way. In this treatment, no discernable effect of added mutant DNA was observed. This indicated that both induction of PL operon expression and mutant DNA addition were required for the enhanced survival. Without wishing to be bound by a single explanation of the observed effects, it is believed that the expressed lambda functions allowed for efficient recombination of this short linear mutant DNA with the chromosomal *galK* gene.

Colonies surviving 2-deoxygalactose treatment were screened for their Gal phenotype on indicator plates, and all tested had the Gal<sup>-</sup> phenotype expected for a *galK* amber mutant. To test specifically whether the *galK* amber mutation was present, four independent Gal<sup>-</sup> colonies were tested by transducing cultures of each with a *lambdaimm21* phage that carries the tRNA<sup>tyr</sup> suppressor allele *supF*. The four mutants tested were suppressed to a Gal<sup>+</sup> phenotype. Finally, the presence of the amber mutation in *galK* was verified by PCR amplification and sequence analysis of the *galK* gene segment from the chromosome.

This example demonstrates that controlled expression of lambda recombination genes from a defective lambda prophage promotes surprisingly efficient homologous recombination in bacterial cells, even with short linear segments of DNA having very short homology arms.

Importantly, the high recombination frequency indicates that recombinants were identified without the need to apply positive or negative selection methods. DNA hybridization probes are thus designed to detect point mutations, deletions, insertions or

other modifications of cellular DNA. Standard colony hybridization or in situ hybridization techniques can be used to detect cells in which recombination has occurred. Alternatively, enrichment methods for mutation detection are used, particularly for detecting point mutations. Such enrichment methods are described in Gocke et al., Annals of the New York Academy of Sciences 906: 31-38, 2000, which is herein incorporated by reference in its entirety. One example of a suitable enrichment is the mismatch amplification mutation assay described by Cha et al., PCR Applications and Methods 2: 14-20, 1992, herein incorporated by reference in its entirety.

The fact that the DNA fragments with short homology arms are able to recombine in vivo opens a vast array of new possibilities for generating recombinant DNA. Several steps normally involved in generating recombinant DNA molecules are eliminated. Restriction enzyme digests are not required to generate DNA fragments, and DNA ligase reactions are not required to join different DNA fragments at novel junctions. The cell generates the completed recombinant precisely joined through homologous recombination.

The efficiency of recombination approaches 0.1% of surviving cells from a standard electroporation. At this efficiency, unselected colonies could be screened for recombinant DNA using colony hybridization techniques, eliminating the need for selection steps. Thus, this recombination protocol makes the bacterial chromosome and plasmid DNA amenable to almost any type of desired change. This includes directed mutagenesis of a gene, a gene segment, or even a base.

#### EXAMPLE 4

##### Preparation of linear DNA cassettes greater than 1000 bp in length

Standard Polymerase Chain Reaction (PCR) conditions were used to amplify linear DNA fragments with the Expand™ High Fidelity PCR system of Boehringer Mannheim. The chloramphenicol resistant ( $Cm^R$ ) cassette *cat* was amplified from pPCR-Script Cam (Stratagene) with primers 5'TGTGACGGAAGATCACTTCG (SEQ ID NO:1) and 5'ACCAGCAATAGACATAAGCG (SEQ ID NO:2). The tetracycline



resistant (Tet<sup>R</sup>) cassette *tet* was amplified from *Tn10* with primers  
5'CTCTTGGGTTATCAAGAGGG (SEQ ID NO:3) and  
5'ACTCGACATCTTGGTTACCG (SEQ ID NO:4). The ampicillin resistant (Ap<sup>R</sup>)  
cassette *amp* was amplified from pBluescript (Stratagene) with primers  
5 5'CATTCAAATATGTATCCGCTC (SEQ ID NO:5) and  
5'AGAGTTGGTAGCTCTTGATC (SEQ ID NO:6). The kanamycin resistant cassette  
*kan* was amplified from *Tn5* with primers 5'TATGGACAGCAAGCGAACCG (SEQ ID  
NO:7) and 5'TCAGAAGAACTCGTCAAGAAG (SEQ ID NO:8). PCR products were  
purified using Qiagen PCR purification kits and concentrated if necessary by ethanol  
10 precipitation. The amplified linear DNAs were suspended in sterile water or TE buffer  
(10mM Tris-Cl pH7.5; 1mM EDTA) and quantified by spectroscopy. DNA in water  
was stored at -20°C. The inventors avoided PCR product purification schemes from  
gels in which the DNA is subject to ultraviolet irradiation.

In order to design primers for amplification of a recombination cassette,  
15 recombinant oligonucleotides were chemically synthesized with the 5' 30 to 50  
nucleotides identical to sequences at the target nucleic acid sequence, and with the 3' 20  
nucleotides homologous to the ends of the cassette to be introduced. A cassette is  
generated by PCR that is flanked by the 30 to 50 base homologies present at the target.

Cells carrying the target DNA either on the chromosome or on a plasmid are  
20 induced for Exo, Beta and Gam function. These cells are made competent for  
electroporation and mixed with the amplified cassette. Following electroporation,  
recombination occurs between the homologous sequences on the linear cassette and the  
target replaces the target segment with the cassette.

The 50 nt *galK* homology segments (rectangles) used for the experiment  
25 described in Table 3 are:  
5'GTTTGC GCGCAGTCAGCGATATCCATTTTCGCGAATCCGGAGTGTAAGAA  
(SEQ ID NO: 9) and  
5'TTCATATTGTTTCAGCGACAGCTTGCTGTACGGCAGGCACCAGCTCTTCCG  
(SEQ ID NO:10)

In one embodiment, the cassette is a drug resistance marker but can be any DNA if the target sequence in the subsequent steps can be counter-selected. The transcription of the marker cassette has been oriented arbitrarily in the same direction as the target region being replaced. The primers contain two parts: a 5' end homologous to flanking regions of the target DNA, and a 3' end that primes the cassette DNA for replication. The PCR using these primers and a DNA template containing the marker cassette generates a linear DNA product with the cassette flanked by target homology. Note that if transformation with the template DNA will generate the selected phenotype (for example, the template is a plasmid), the template is then eliminated. Plasmid template DNA can be destroyed by treatment with DpnI following the PCR; DpnI cuts methylated GATC template DNA leaving the newly replicated unmethylated DNA intact. Once a linear cassette has been generated, it can be stored and used as the template for subsequent PCRs.

### EXAMPLE 5

#### Gene replacement by targeted homologous recombination

Having demonstrated in Example 3 that 70-bp linear DNA can direct mutations to a specific target, a synthetic DNA having 50-bp *galK* DNA segments flanking the *cat* (chloramphenicol resistance, or  $\text{Cm}^R$ ) cassette was constructed for targeting a *galK* gene replacement by *cat*.

The linear *cat* cassette with flanking *galK* DNA was made by PCR using chemically synthesized primers, as described in Example 4.

Data from these experiments are presented in Table 3. DY330 competent cells were electroporated with 100 ng of the *cat* cassette targeted to replace either *galK* (*galK*  $\diamond$  *cat*) or prophage genes *cIII kil gam* (*cIII kil gam*  $\diamond$  *cat*; see FIG. 1 for map of prophage genes). Total recombinants per electroporation are shown in the rightmost column, "CmR recombinants." The *cat* cassette was transferred by electroporation into *galK*<sup>+</sup> cells which had either been heat-induced for PL operon expression (15 min temperature shift to 42°C, as indicated by "15" in the center column of Table 3) or not

induced (maintained at 32°C; indicated by "0" in the center column of Table 3). See also Example 2 for description of induction and other methods. After electroporation, Cm<sup>R</sup> recombinants were selected at 32°C, and then quantified.

- As shown in Table 3, Cm<sup>R</sup> colonies were only found in the heat-induced culture.
- 5 All 50 Cm<sup>R</sup> colonies tested had a Gal<sup>-</sup> phenotype on MacConkey galactose indicator agar, indicating the presence of the *galK*<>*cat* replacement. The symbol <> indicates a replacement generated by homologous recombination techniques, for example, *galK*<>*cat* indicates that the bacterial *galK* gene is replaced by *cat* using homologous recombination techniques.

10

Table 3

Target Site*	42°C, min	Cm <sup>R</sup> Recombinants
<i>GalK</i>	0	<1
<i>GalK</i>	15	2.5 x 10 <sup>4</sup>
<i>CIII kil gam</i>	0	<1
<i>CIII kil gam</i>	15	5.0 x 10 <sup>4</sup>

- In similar experiments using the same 50-bp homologous arms, the *galK* has been exchanged for *kan*, *amp*, and *tet* cassettes by selecting for Km<sup>R</sup>, Amp<sup>R</sup>, and Tet<sup>R</sup>, creating *galK*<>*kan*, *galK*<>*amp*, and *galK*<>*tet* replacements, respectively.
- 15

- To test whether this approach also works at other positions on the bacterial chromosome, a linear *cat* cassette was created flanked by 50-bp DNA segments found immediately upstream and downstream of the *rnc* gene encoding RNaseIII. The *rnc* gene is thought to be non-essential (Takiff et al., Journal of Bacteriology 171: 2581-2590, 1989); therefore, it was tested whether an exact substitution of the *cat* coding region for the *rnc* coding region (from AUG to codon 224) could be made using the recombination techniques herein described. In this construct, *cat* is transcribed from the *rnc* promoter, and the 5' primer used to generate *cat* started at the *cat* initiation codon.
- 20

Following procedures used for *galK* as described in Examples 2 and 3, Cm<sup>R</sup> colonies were found, but only in the induced culture. The Cm<sup>R</sup> colonies tested had a Rnc mutant phenotype, as described in Example 2.

Two other *rnc*<>*cat* recombinants were made. One replaced sequence from the  
5 AUG start to codon 126 of *rnc*, and the other from the AUG start to codon 192 of *rnc*.  
These two recombinants generate *cat::rnc* gene fusions with an *rnc* mutant phenotype.  
Different sets of primers were chosen to detect unambiguously the wild-type and/or  
recombinant alleles. This PCR procedure follows guidelines set forth by yeast  
researchers in characterizing chromosomal replacements in yeast (Winzeler et al.,  
10 Science 285: 901-906, 1989). The PCR analysis of the recombinants verified the loss of  
the *rnc*<sup>+</sup> gene and the predicted structures of the three *rnc*<>*cat* gene replacements.

In this example, several different genes on the bacterial chromosome and on  
plasmids have been substituted with drug resistance markers. However, it is also  
possible to create recombinants in which the desired product does not include a  
15 selectable marker. Genes have been fused to cassettes encoding specific tags such as  
the green fluorescent protein. Fusion tags can be placed precisely in the gene to be  
modified, for example, by any of the following strategies. In one, the unselected  
cassette is joined to a selectable drug marker, and both are recombined into the chosen  
location selecting for drug resistance. In another, the cassette is recombined into its  
20 location by substituting it for a negative selection marker like *sacB* (Blomfield et al.,  
Mol Microbiol 5, 1447-57, 1991). This strategy permits cloning of any DNA. In a third  
strategy, the recombinants are screened non-selectively by DNA hybridization with  
probes specific to the cassette.

In these experiments, the desired recombination product was usually obtained.  
25 However, some recombination products were unexpected. In two cases, an attempt was  
made to knock-out essential genes, and surprisingly it was possible to select a few rare  
recombinants. These turned out to be diploid for the region of the targeted gene, since  
they carried the wild type and the mutant copy of the gene as determined by PCR  
analysis. Rare diploid regions of the bacterial chromosome are known to occur

spontaneously in growing cells at a frequency of about 0.1% (Haack and Roth, Genetics 14, 1245-1252, 1995). Because an essential gene was targeted, these rare diploids were selected. This was only possible because the recombination is so efficient.

This example demonstrates that the methods of this invention can be used to promote efficient gene replacement by homologous recombination. The gene replacements were made throughout the bacterial chromosome. In Example 7, it is shown that the methods can be used to modify extrachromosomal nucleic acids, such as plasmids, bacterial artificial chromosomes, cosmids, phagemids, and the like.

#### EXAMPLE 6

##### Induction time, DNA amount, and homology arm length affect targeting efficiency

*Induction time.* FIG. 3 shows the effect of induction time on recombination. The strain DY330 was grown at 32°C to OD<sub>600</sub>=0.4 to 0.8, heat-induced at 42°C for the times indicated and then made electrocompetent (see Example 2 for description of methods). A linear *cat* cassette (10 ng) was used to target the *cIII kil gam* genes of the prophage. Total Cm<sup>R</sup> recombinants were plotted vs. the time of induction.

Induction of PL operon expression for only five minutes enhanced recombination activity. FIG. 3 reveals that by 7.5 min of heat induction a maximum efficiency is reached. This maximum level is maintained for induction times from 7.5 to 17.5 min with some reduction occurring for times longer than 17.5 min. Expression of the PL operon for longer than 60 min causes cell death.

Cells harboring the defective lambda prophage may be grown at temperatures other than 32 °C. In general, it is undesirable to grow cells at temperatures >37 °C, because such temperatures lead to partial inactivation of the cI repressor, and leaky expression from the PL promoter. In general, it is also undesirable to grow cells at temperatures below 20°C, because of slow growth. One skilled in the art would also recognize that there is a considerable degree of latitude with regard to time and

temperature of induction. For example, expression of lambda recombination genes from the PL operon could be induced at temperatures as low as 38 °C, generally allowing for longer times of induction. The limit for protein expression in *E. coli* is about 45°C.

5        *Donor DNA amount.* FIG. 4 shows the effect of amount of the linear DNA cassette on recombination. The strain DY330 was grown at 32°C to OD<sub>600</sub>=0.4 to 0.8, induced at 42°C for 15 min and then made electrocompetent. Different amounts (1, 10, 100, 300, 1,000 ng) of a linear *cat* cassette (1 kbp in length) were used to target the *cIII* *kil gam* genes of the prophage. Total Cm<sup>R</sup> recombinants were plotted vs. the DNA  
10        amount at 42°C.

Fig. 4 shows that targeting efficiency increased in a near linear relationship with increasing concentration of donor DNA in the range from 10<sup>8</sup> (1 ng) to 10<sup>10</sup> (100 ng) molecules per electroporation. A saturating level of linear DNA is reached at 3x10<sup>10</sup> molecules yielding 7.5x10<sup>4</sup> recombinants per ~2x10<sup>8</sup> cells electroporated. Thus, the  
15        methods of this invention may be practiced over a broad range of oligonucleotide concentrations.

*Homology length.* FIG 5 shows the effect of homologous arm length on recombination. The strains DY330 (*recA*+) (filled circles) and DY331 (*recA*-) (open circles) were grown at 32°C, induced at 42°C for 15 min and then made  
20        electrocompetent. A linear *cat* cassette (100 ng) was used to target the *cIII kil gam* genes of the prophage. The homologous arm length of the cassette was varied from 0 to 1,000 bp. The primers containing the 0 to 50 bp homologies were chemically synthesized as described (Fig. 2). The cassette containing 1,000 bp homologous arms was made by PCR using primers 1,000 bp away on each side of an existing (*cIII kil*  
25        *gam*)<>*cat* disruption in the cell. Total Cm<sup>R</sup> recombinants were plotted vs. the homologous arm length.

Several pairs of primers were made to amplify the *cat* cassette for targeting the chromosome and designed each pair with a different length of flanking homology. The length of the homology segment on the primers varied by increments of 10 bases from

10 to 50 bases. A nested set of linear *cat* cassettes was made with the primers. Another linear *cat* cassette was constructed flanked by 1,000 bp of homology. This set of linear DNAs was then tested for recombinational targeting efficiency as shown in FIG. 5. No recombinants were found with 10 bp of homology and less than ten recombinants were found in each of three experiments with 20 bp of homology. From 20 bp to 40 bp of homology, homologous recombination increased by four orders of magnitude. From 40 bp to 1,000 bp of homology, recombination increased only 10-fold.

These data indicate that the methods herein disclosed may be practiced with surprisingly short homology arms, as few as 20-40 residues. However, homology arms of 30 or greater residues increase efficiency.

#### EXAMPLE 7

##### Gene replacement on plasmids and BACs: in vivo cloning

To determine if this method could be used to modify plasmid DNA, the procedures described in Examples 2-3 were followed to modify plasmid pGB2, a derivative of pSC101 (Bernardi et al., Nucleic Acids Res. 12: 9415-9426, 1984). A *cat* cassette was synthesized in vitro and recombined in vivo with pGB2 to replace the spectinomycin resistance gene with *cat* conferring Cm<sup>R</sup> on the cell carrying the recombinant plasmid.

The same experiment, performed on pBR322 derivatives, generated recombinants, but they were joined in tandem to non-recombinant plasmids as dimers and higher multimers. Induction of Gam expression from our prophage inactivates RecBCD nuclease. In the absence of RecBCD, pBR322 derivatives replicate by a rolling circle mode (Feiss et al, Gene 17: 123-130, 1982), and the plasmid converts from monomers to multimers. This is specific for pBR322-type replicons as the pGB2 type did not form multimers.

To generate simple recombinants of pBR322 derivatives, the protocol was modified by coelectroporating the *recA*<sup>-</sup> strain DY331 with circular plasmid DNA (0.1

ng) and a linear drug cassette. Recombinant plasmid monomers were readily selected and isolated.

In addition to plasmids, the method is also suitable for targeting genes on bacterial artificial chromosomes, phagemid artificial chromosomes, yeast artificial  
5 chromosomes, cosmids, and the like. Homologous recombination between target nucleic acid sequences on BACs and synthetic oligonucleotides (such as ds DNA fragments or PCR fragments) is carried out in bacterial cells bearing the defective lambda prophage shown in FIG. 1 and described in Example 1. Synthetic  
oligonucleotides (such as short annealed dsDNA fragments or PCR fragments) are  
10 electroporated into bacterial cells as described in Example 2. Analysis for successful recombination events is by selective PCR amplification using specific primers for the introduced sequence, or by selective amplification approaches such as the mismatch  
amplification mutation assay described by Cha et al., PCR Applications and Methods 2:  
14-20, 1992. or by direct hybridization using specific probes. Using this approach  
15 recombination frequencies of up to 1:500 have been observed, regardless of the strand targeted. Thus, this system is extremely useful in manipulation of and rapid screening for recombinants in BAC vectors. The unexpectedly high efficiencies eliminate the need for introducing selectable markers, or modifying such markers on the BAC.

## EXAMPLE 8

20

### Requirement of RecA for targeted recombination

The requirement for RecA were tested in targeted homologous recombination by repeating the experiment described in Fig. 5 but using a *recA*<sup>-</sup> strain. Surprisingly, recombination efficiency was depressed only about 10-fold in the *recA*<sup>-</sup> mutant for the  
25 arm lengths tested (Fig. 5). Thus, RecA function is not required, and *recA*<sup>-</sup> strains mediate efficient homologous recombination with linear DNA fragments having homology arms of 30 bp or greater.

This result was unexpected. The  $\lambda$  recombination system is known to function in cells lacking the bacterial RecA function (Brooks and Clark, J. Virol. 1: 283-293,



1967). However, the recombination in *recA* mutants under conditions used by others is reduced more than 50-fold relative to levels in *recA*<sup>+</sup> cells (Stahl et al., Genetics 77: 395-408, 1974; Murphy, J. Bacteriol. 180: 2063-2071, 1998).

### EXAMPLE 9

#### Lambda genes promote targeted recombination of dsDNA in wild type *E. coli*

To determine which lambda genes promote targeted recombination of dsDNA, a set of replacement deletions were generated in the PL operon of the prophage using *cat* and *amp* cassettes. In the center column of Table 4, the parentheses indicate the deletions made by the recombination event within the prophage (refer to Fig. 1 for a linear map of the prophage genes). Each of these newly made deletions was verified structurally by PCR analysis and tested for targeted recombination of a *tet* gene cassette into *galK*.

Electrocompetent cells from strains indicated in Table 4 were heat-induced for 15 minutes at 42 °C, and electroporated with 10 ng of linear *galK*◊*tet*. The results are presented as total number of Tet<sup>R</sup> recombinants per electroporation in the right hand column of Table 4.

Table 4

Strain	Prophage	Recombinant
DY330	wild-type	4,100
DY392	( <i>hin-int</i> )◊ <i>amp</i>	2,000
DY351	( <i>sieB-kil</i> )◊ <i>cat</i>	4,400
DY386	( <i>hin-int</i> )◊ <i>amp</i>	1,650
	( <i>sieB-kil</i> )◊ <i>cat</i>	
DY349	( <i>gam</i> )◊ <i>cat</i>	0

DY360	( <i>bet</i> ) $\diamond$ <i>cat</i>	0
DY359	( <i>exo</i> ) $\diamond$ <i>cat</i>	0

Table 4 shows that only *exo*, *bet*, and *gam* deletions affected *galK* $\diamond$ *tet* targeted recombination. Deletion of any one of these three genes eliminated the recombination, whereas deletion of all other genes in the pL operon had little if any effect.

5 To show that the *gam* $\diamond$ *cat* substitution was not polar on *bet* and *exo*, the *gam* gene was expressed in trans and shown to complement the defect. Thus, although the entire PL operon was used in studies described here, only *exo bet* and *gam* functions are needed for recombination with double-stranded DNA cassettes >100 –200 bp made by PCR.

10

### EXAMPLE 10

#### Efficient Recombination with Single-Strand DNA

To evaluate recombination between exogenous single-strand DNA and the E. coli chromosome, the experiments described in this Example were performed. In addition, the experiments also address the role of the bacterial *recA* recombination genes in mediating the observed effects.

15 Expression from the PL operon was heat-induced, cells were made electroporation-competent, and 70-mer oligonucleotides were electroporated into cells, all as described in Example 2. Strains HME9 and HME10 both harbor the *galK* amber mutation, and are therefore Gal<sup>-</sup> in phenotype. They differ from each other in that HME9 is *recA*<sup>+</sup>, whereas HME10 is *recA*<sup>-</sup>. Thus in this experiment recombination efficiencies in *recA*<sup>+</sup> and *recA*<sup>-</sup> cells are compared.

20 The 70-mer single-stranded oligonucleotides used in this experiment were designed to restore wild-type *galK* gene activity (hereinafter *galK*<sup>+</sup>) upon successful recombination, thereby producing a Gal<sup>+</sup> phenotype (i.e., ability to grow on minimal media with galactose as the sole carbon source). The 70-mer corresponding to the

transcriptional non-template DNA strand of galK was

5'AAGTCGCGGTCGGAACCGTATTGCAGCAGCTTTATCATCTGCCGCTGGAC  
GGCGCACAAATCGCGCTTAA (SEQ ID NO:11).

70-mer single-stranded DNA of either SEQ ID NO:11 or its complement was  
5 electroporated into cells. Alternatively, the two 70-mers were first annealed to each  
other and then electroporated into cells as double strand DNA. A successful  
recombination event was identified by restoration of the Gal<sup>+</sup> phenotype. Table 5  
indicates the number of galK<sup>+</sup> recombinants per viable cell,  $\times 10^{-4}$ .

Table 5 presents the number of recombinants observed  $\times 10^{-4}$  after  
10 electroporation of the HME9 or HME10 strains with DNA in the indicated forms.  
Efficient recombination was observed with double strand DNA, similar to that  
previously described in Examples 3 and 5. Surprisingly, single-strand DNA was about  
equally or even more efficient than double-stranded DNA in producing homologous  
recombination, regardless of strand used. In this experiment, recombination efficiencies  
15 for the double-stranded DNA was about 1 in 3800 cells ( $2.6 - 2.7 \times 10^{-4}$ ), whereas  
recombination efficiency for single-stranded counterclockwise DNA was 3-to 7-fold  
higher. In other strains (e.g DY374) recombination of single strand linear DNA as  
frequent as one nadA<sup>+</sup> recombinant per 20 viable cells has been observed. In addition,  
efficient recombination was observed in both recA<sup>+</sup> and recA<sup>-</sup> cells, establishing that  
20 the recombination events did not require the bacterial recA gene products.

Table 5

Strain Used	DNA used in electroporation		
	dsDNA (about 1 µg)	cc-ssDNA (about 0.7 µg)	cw-ssDNA (about 0.6 µg)
5 HME9 <i>galK<sup>am</sup>λCI857Δ(cro-bio)</i>	2.6	18.5	2.2
10 HME10 <i>galK<sup>am</sup>λCI857Δ(cro-bio) recA<sup>-</sup></i>	2.7	9	0.35

cc=clockwise (strand with the 5' to 3' orientation relative to transcription)

cw=counterclockwise (strand with the 3' to 5' orientation)

15

### EXAMPLE 11

#### Efficient Generation of Large Deletions by Recombination with Single-strand DNA

20 Example 10 demonstrates efficient lambda-mediated recombination using ssDNA to generate a single base change in the *E. coli* chromosome. This example demonstrates similar high efficiency when the approach is used to generate large deletions.

25 Using the methods described in Examples 2, 3, and 10, lambda operon expression was heat-induced, cells were rendered electroporation-competent, and 70-mer single-stranded DNA of either SEQ ID NO:11 or its complement was electroporated into cells. The DNA was electroporated into two strains: one containing the *galK* amber mutation, and one in which the *galK* gene was interrupted by a *cat-sacB* cassette precisely inserted at the position of the amber mutation in *galK*. The strains

were otherwise genetically identical. A successful recombination event was identified by restoration of the Gal<sup>+</sup> phenotype. Table 6 indicates the number of galK<sup>+</sup> recombinants per viable cell,  $\times 10^{-4}$ .

The data in Table 6 demonstrate that recombination efficiency is similar using the same oligonucleotide, regardless of whether lambda-mediated ssDNA recombination is being used to generate a single base change or a large deletion, removal of the 3264 bp *cat-sacB* cassette. In both cases, the method is highly efficient in generating recombinants.

Table 6

Strain Used	DNA used in electroporation	
	cc-ssDNA (200 ng)	cw-ssDNA (200 ng)
HME6 galK <sup>am</sup> λCI857Δ(cro-bio)	15	3
HME31 galK<catsacB λCI857Δ(cro-bio)	10	0.5

The length that is reasonable to synthesize chemically limits the length of single strand oligonucleotides used for recombination. It is demonstrated herein that two oligonucleotides that have a complementary overlap region at their 3' ends, when co-electroporated into DY411 cells, can anneal and generate recombinants with chromosomal or extrachromosomal DNA (Figure 14). This recombination requires the induction of the PL operon and the Gam, Beta, and Exo functions. A *galK* mutation in which the *kan* cassette was placed in *galK* in a way to delete 34 bp of the *galK* gene was created by recombineering. Two oligos were synthesized that were 70 bases long, with

34 bases of the deleted region at their 3' ends, that were complementary and can act to anneal the two oligos together. The 5' end of each oligo contained 36 bases of homology to each side of the 34bp deletion caused by *kan*. Each oligo alone cannot generate *gal*<sup>+</sup> recombinants but mixed together they generated up to 10<sup>5</sup> recombinants per 10<sup>8</sup> cells electroporated. Oligos with the same sequence but shortened from their 3' end to overlap by only 2 bases did not yield recombinants. However, overlaps of 10 bases or more generated recombinants. Preannealing was not required and the two oligos can be mixed and used directly for electroporation.

If the ends of the overlaps are filled in by DNA polymerase a 104 bp duplex is generated. This dsDNA generates only slightly more recombinants than the DNA with 10 to 34 base overhangs. Thus, multiple overlapping (by >10 bases) oligonucleotides of any even number can be used to yield recombinants, in which the most outside oligonucleotides have 5' overlaps. The end oligonucleotides also have 30-50 bases of homology to the targeted region. The use of multiple overlapping oligonucleotides allows production of long recombination substrates without use of PCR. The central oligonucleotide(s) can be any cassette envisioned to be used for dsDNA recombination. This recombination with overlapping oligonucleotides having outside 5' overhangs is most efficient with Exo, Beta, and Gam, but can be recombined by Beta alone (without Exo and Gam) in the cell. This greatly simplifies the recombination procedure (as only Beta is required). Although the 104 bp duplex DNA recombines more efficiently if Exo, Beta and Gam are present, recombination also occurs in the absence of Gam, albeit at a lower efficiency (the duplex requires both Exo and Beta for recombination).

Similar overlapping synthetic oligonucleotides can be generated with 3' overhangs of 34 bases that can be coelectroporated into cells. These are also recombined into targets defined by homology at the ends. Again, only Beta is required for this recombination. In this case, Exo is not required, and further Exo does not stimulate recombination. In one embodiment, multiple oligonucleotides can be overlapped as above to span longer distance. As long as the outermost oligonucleotides have 3' overhangs, recombination will be Exo independent. The efficiencies of the

present system allows the detection of recombinants in this case.

Examples 10 and 11, taken together, document that the methods disclosed herein can be practiced with ssDNA oligonucleotides. This surprising result enables high efficiency homologous recombination with synthetic DNA of single or double strandedness.

The present system allows the limit of synthetic oligonucleotide size to be increased dramatically by overlapping oligonucleotides. In addition, the system allows recombination of these DNAs to be carried out with Beta alone, or Exo with Beta but without Gam. Recombination without the requirement for Gam is important because Gam is the toxic function that was a limiting factor in the previously described methods. As the present system requires only Beta, a constitutive promoter can be utilized.

## EXAMPLE 12

### Effect of ssDNA Length on Recombination Efficiency

In Examples 10 and 11, lambda-mediated recombination was used to efficiently incorporate 70-mer ssDNAs into the *E. coli* chromosome. In this example, the effect of oligonucleotide length on recombination efficiency was investigated.

Using the methods described in Examples 2, 3, 5, 10 and 11, lambda operon expression was heat-induced, cells were rendered electroporation-competent, and ssDNA oligonucleotides (200 ng each) were electroporated into *E. coli* HME9 strain cells. The electroporated ssDNA oligonucleotides included the 70-mer of SEQ ID NO:11, a 60-mer constructed by removing the last 5 nucleotides from both the 5' and 3' ends of SEQ ID NO:11, and a 50-mer, 40-mer, 30-mer, or 20-mer constructed by removing the last 10, 15, 20, or 25 nucleotides, respectively, from both the 5' and 3' ends of SEQ ID NO:11. As in example 10, the ssDNA oligonucleotides used in this experiment were all designed to restore the galK+ gene upon successful recombination, thereby conferring upon the cell a Gal+ phenotype. Table 7 indicates the number of galK+ recombinants per viable cell,  $\times 10^{-4}$ .

Table 7

		Oligonucleotide length						
		0	20	30	40	50	60	70
5	Efficiency ( $\times 10^{-4}$ )	0.004	0.01	0.47	4	4	6	22

As the data in Table 7 demonstrate, recombination efficiency increases with increasing ssDNA length. Recombination efficiency was low when the ssDNA used was a 20-mer, but increased considerably with a 30-mer. Efficiency was near optimal with a 40-mer, and increased to 1 in 450 viable cells with a 70-mer. Hence, specific examples of the invention use single-stranded DNA molecules at least about 40 nucleotides in length.

Without wishing to be bound by a single explanation of the observed effects, the inventors currently believe that observed length-efficiency relationship may reflect published data indicating that lambda Beta protein binds stably to DNA sequences of 36 bases or longer, but does not bind as well to shorter oligonucleotides (Mythili et al., Gene 182: 81-87, 1996).

### EXAMPLE 13

#### Lambda Beta Protein Mediates Efficient Recombination with ssDNA

To determine whether lambda Beta protein was sufficient to mediate recombination between exogenous ssDNA and the *E. coli* chromosome, the efficiency of recombination was investigated in a strain that expressed lambda Beta, but not Exo or Gam.

For these experiments, the HME43 strain was used. Its genotype is identical to the HME6 strain, except that the lambda prophage contains additional genetic deletions,



from *int* through *exo* and from *gam* through *N* (see FIG. 1). In addition, the *cat* gene conferring the CmR phenotype is inserted between *attL* and *bet*.

Using the methods described in Examples 2, 3, and 10-12, expression of the modified lambda operon was heat-induced, cells were rendered electroporation-competent, and 70-mer ssDNA of SEQ ID NO:11 (200 ng) was electroporated into cells. Using this procedure, the HME43 strain expresses lambda Beta protein, but does not express *gam*, *exo*, or any other prophage encoded genes. A successful recombination event was identified by restoration of the Gal<sup>+</sup> phenotype. Table 8 indicates the number of galK<sup>+</sup> recombinants per viable cell, x 10<sup>-4</sup>.

Table 8

10

Strain	Prophage Modifications	Recombination Efficiency (x 10 <sup>-4</sup> )
HME43	( <i>int-exo</i> )>cat, ( <i>gam-N</i> )>Δ	7.7

15

In contrast to Example 9 using PCR-generated double-stranded DNA, the data presented in this example establish that lambda Beta *alone* is sufficient to mediate efficient recombination between ssDNA and the *E. coli* chromosome. Moreover, two or more overlapping oligonucleotides may be used, if they have a 3' overhang and more than about 10 bp of overlap. Overlapping oligonucleotides with a 5' overhang also promote homologous recombination with Beta alone. However, for 5' overhangs, *exo* and *gam* (or a similar exonuclease and RecBCD-inhibition function) appear to enhance maximal efficiency.

20

A modification of the method is to place DNA encoding other ssDNA binding polypeptides under control of the PL promoter. For example, the strain HME43 is further modified to delete *bet* and insert DNA encoding P22 Erf, RecT, or Rad52. Expression of the ssDNA binding polypeptide is induced by temperature shift as it is for

25

induction of lambda *bet* expression. Exo and Gam, or proteins with similar function, may also be placed under control of the PL promoter. Moreover, other inducible or constitutive promoters may be used.

5

#### EXAMPLE 14

##### Ex Vivo Combination of ssDNA With Lambda Beta Mediates Efficient Homologous Recombination

I wanted you to know that this experiment has not worked yet. It is still a n  
experiment that has not yet been completed.

10

Single-strand DNA can be combined with lambda Beta protein prior to electroporation into cells, and mediated efficient recombination between the ssDNA and the host DNA.

Lambda Beta proteins may be prepared by techniques known in the art  
(Karakousis et al., J. Mol. Biol. 276: 721-731, 1998), and preincubated at 37 °C with  
single-strand oligonucleotides of 20-mer or greater length. In this example, ssDNA  
oligonucleotides of SEQ ID NO:11, a 60-mer constructed by removing the last 5  
nucleotides from both the 5' and 3' ends of SEQ ID NO:11, and a 50-mer constructed  
by removing the last 10 nucleotides from both the 5' and 3' ends of SEQ ID NO:11  
were used. Typically, lambda Beta protein concentration is about 2.5 µM and DNA  
concentration about 5 µM, but the method is effective with a broad range of protein and  
DNA concentrations (for example, from 0.1 µM to 10 mM protein, and 0.01 µM to 10  
mM ssDNA). Alternatively, the Beta protein and ssDNA can be coelectroporated into  
cells without premixture or preincubation.

20

The DNA and protein is electroporated into *E. coli* using methods described in  
Examples 2 and 3. In this example HME 43 strain is used, but numerous other strains  
are suitable. Expression of the modified lambda operon in one set of cells is heat-  
induced, and a second set of cells is maintained at 32 °C. Both sets of cells are rendered

25

electroporation-competent, and 70-mer ssDNA of SEQ ID NO:11 (200 ng) is electroporated into both heat-induced and uninduced cells. Using this procedure, the HME43 strain expresses lambda Beta protein upon temperature shift to 42°C, but does not express Beta from *bet* or any other prophage-encoded genes in the absence of a temperature shift. A successful recombination event is identified by restoration of the Gal<sup>+</sup> phenotype.

In this experiment, high efficiency recombination is observed in both heat-induced and uninduced cells. Moreover, it is believed that approximately equally high efficiency recombination is observed when these techniques are followed in *E. coli* strains that contain no lambda prophage genes.

This approach may be modified by substituting other ssDNA binding polypeptides for lambda Beta, such as p22 Erf, RecT and Rad52. The target nucleic acid sequence may be on the bacterial chromosome, or on exogenous DNA such as a bacterial artificial chromosome, phagemid artificial chromosome, plasmid, cosmid, or the like. Moreover, there is no particular requirement for a specific bacterial species; these single-strand DNA binding polypeptides will mediate efficient recombination in a broad range of bacteria. Indeed, these polypeptides will mediate efficient recombination in eukaryotic cells as well, as in Example 15.

## EXAMPLE 15

### Lambda Beta Protein Mediates Efficient Homologous Recombination in Eukaryotic Cells

The ex vivo approach described in Example 14 may be used to target genes in eukaryotic cells for homologous recombination. In eukaryotic cells, transfection of the ssDNA with lambda Beta protein may be accomplished by electroporation as in Examples 2, 3 and 14, or by the methods of Chang et al., *Biochimica et Biophysica Acta*, 153-160 (1992), Keating and Toneguzzo, *Bone Marrow Purging and Processing*, 491-498 (1990), or other electroporation protocols known in the art. In addition, a

variety of means for macromolecular transfer methods are known to the art, including calcium phosphate-DNA co-precipitation (Ausubel et al.), DEAE-dextran-mediated transfection (Matthews et al, *Experimental Hematology*. 21: 697-702, 1993) polybrene-mediated transfection (Costello et al., *Gene Therapy* 7: 596-604, 2000), microinjection  
5 (Davis et al., *Blood* 95: 437-44, 2000), liposome fusion and lipofection (Veit et al., *Cardiovascular Research*. 43: 808-22, 1999), protoplast fusion (Schaffner, *Proc. Natl. Acad. Sci. USA* 77: 2163, 1980), inactivated adenovirus-mediated transfer (Wagner et al., *Proc Natl Acad Sci USA* 89:6099-6103, 1992), hemagglutinin virus of Japan- (HVJ)-mediated transfer (Morishita et al., *Journal of Clinical Investigation* 93: 1458-1464,  
10 1994), biolistics (particle bombardment) and the like. Any such macromolecular transfer approach is suitable. Design of dsDNA molecules for facilitating homologous recombination with eukaryotic genes is well known in the art (for example, as described in Mansour, *Nature* 336: 348-352, 1988; Shesely, *PNAS* 88: 4294-4298, 1991; Capecchi, M. R., *Trends in Genetics* 5: 70-76, 1989; U.S. Patent No. 6,063,630).

15 Cells to be transfected with exogenous DNA are combined with a DNA construct comprising the exogenous DNA, targeting DNA sequences and, optionally, DNA encoding one or more selectable markers. The resulting combination is treated in such a manner that the DNA construct enters the cells. This is accomplished by  
20 introducing DNA into vertebrate cells. Once in the cell, the exogenous ssDNA is integrated into cellular DNA by homologous recombination between DNA sequences in the DNA construct and DNA sequences in the cellular DNA.

For example, the target nucleic acid is the beta-globin gene in hematopoietic stem cells from a patient with sickle cell anemia (Beutler, *Disorders of Hemoglobin*,  
25 Ch. 107 in Harrison's *Principles of Internal Medicine*, 14<sup>th</sup> ed. © 1998, herein incorporated by reference). The sickle cell Beta globin gene harbors a point mutation that substitutes a Val for Glu at position six of the polypeptide chain, resulting in an abnormal hemoglobin which is prone to inappropriate polymerization. The methods of this invention can be used to correct the mutation.

Hematopoietic stem cells from a sickle cell patient are isolated, cultured, and expanded ex vivo as is known in the art (Brugger, Seminars in Hematology 37[1 Suppl 2]: 42-49, 2000; Dao et al., Blood 92: 4612-21, 1998; Aglietta et al., Haematologica 83: 824-48, 1998; Emerson, Blood 87: 3082-8, 1996). A 60-mer ssDNA oligonucleotide of  
5 SEQ ID NO:12 (AACAGACACC ATGGTGCACC TGACTCCTGA  
GGAGAAGTCT GCCGTTACTG CCCTGTGGGG) is synthesized and partially purified by standard techniques (Pfleiderer et al., Acta Biochimica Polonica. 43:37-44, 1996; Anderson et al., Applied Biochemistry & Biotechnology 54: 19-42, 1995, herein incorporated by reference).

10 After culture and ex vivo expansion, about  $10^6$  hematopoietic stem cells are suspended in 0.4 mL PBS containing 0.1% glucose, about 10  $\mu$ M purified lambda Beta protein, and about 1  $\mu$ g ssDNA oligonucleotide of SEQ ID NO:12. The cell suspension is electroporated in a 1-mL cuvette at 280V and 250  $\mu$ F with a Gene Pulser (Bio-Rad Laboratories Inc., Hercules, California, USA). Cells are then plated and cultured.  
15 Homologous recombinants harboring the mutation are identified and clonally isolated, further expanded ex vivo, and may be returned to the patient, or cultured for additional in vitro study.

Those skilled in the art will recognize that a broad range of ssDNA and ssDNA binding polypeptide concentrations will be effective, as in Example 14. For example,  
20 both ssDNA and ssDNA binding protein may be present in concentrations ranging from 0.001  $\mu$ M to 100 mM; or from 0.1 mM to 1  $\mu$ M; or from 1  $\mu$ M to 100  $\mu$ M. Oligonucleotide length can be varied in accordance with parameters presented in Example 12. There is no particular upper limit on oligonucleotide length. In addition, two or more oligonucleotides can be included which have complementary 5' ends,  
25 thereby creating 3' overhangs which are effective substrates for ssDNA binding polypeptides such as lambda Beta. In addition, RecT, P22 Erf, Rad52, and other double strand break repair ssDNA binding polypeptides may be substituted for lambda Beta. Culture and electroporation conditions are readily variable without materially reducing

homologous recombination. Moreover, nucleic acid may be introduced into the cell by any suitable macromolecular transfer method.

Other types of stem cells may be used to correct the specific gene defects associated with cells derived from such stem cells. Such other stem cells include  
5 epithelial, liver, lung, muscle, endothelial, mesenchymal, neural and bone stem cells.

Alternatively, certain disease states can be treated by modifying the genome of cells in a way that does not correct a genetic defect per se but provides for the supplementation of the gene product of a defective gene. For example, endothelial cells are preferred as targets for human gene therapy to treat disorders affecting factors  
10 normally present in the systemic circulation. In model studies using both dogs and pigs endothelial cells have been shown to form primary cultures, to be transformable with DNA in culture, and to be capable of expressing a transgene upon re-implantation in arterial grafts into the host organism. Wilson et al., Science, 244: 1344, 1989; Nabel et al., Science, 244: 1342, 1989. Since endothelial cells form an integral part of the graft,  
15 such transformed cells can be used to produce proteins to be secreted into the circulatory system and thus serve as therapeutic agents in the treatment of genetic disorders affecting circulating factors. Examples of such diseases include insulin-deficient diabetes, alpha-1-antitrypsin deficiency, and hemophilia. Epithelial cells, myocytes and hepatocytes are also useful cell types for therapeutic production of  
20 proteins.

The method is also useful for knockout or modification of genes in embryonic stem (ES) cells. Such cells have been manipulated to introduce transgenes. ES cells are obtained from pre-implantation embryos cultured in vitro. Evans et al., Nature, 292, 154-156, 1981; Bradley et al., Nature, 309, 255-258, 1984; Gossler, et al., Proc. Natl.  
25 Acad. Sci. U.S.A., 83, 9065-9069, 1986; Robertson, et al., Nature, 322, 445-448, 1986; U.S. Patent No. 5,464,764. Oligonucleotides designed to target specific gene segments in the ES cell are combined with lambda Beta protein or other ssDNA binding polypeptide and introduced into ES cells by electroporation or other transformation methods. The oligonucleotides may be designed as a series of overlapping segments

with 3' overhangs. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and can contribute to the germ line of the resulting chimeric animal (Jaenisch, Science, 240, 1468-1474, 1988).

5 For example, sequences encoding positive selection marker neomycin resistance gene are synthesized as a series of overlapping 70-mer oligonucleotides, 20 base pairs of overlap and 3' overhangs. The 3' terminal oligonucleotides are designed to insert into the second exon of the mouse hox 1.1 gene as described in U.S. Patent No. 5,464,764. Because the overlapping oligonucleotides combine to encode a promoterless  
10 neomycin resistance gene, only those that successfully incorporate into the targeted mouse hox 1.1 second exon will express the neo gene product and have the neomycin resistance phenotype. The targeting is designed to provide the synthetic neomycin resistance gene with an operable promoter and translation start derived from the mouse hox 1.1 gene. The targeting DNA is also designed so that random incorporations  
15 elsewhere in the ES cell genome are unlikely to be operably linked to any promoter to allow transcription and translation.

The series of overlapping oligonucleotides with 3' overhangs (about 200 nanograms each) are combined with 10  $\mu$ M lambda Beta protein and introduced into ES cells by electroporation using the Promega Biotech X-Cell 2000. Rapidly growing cells  
20 are trypsinized, washed in DMEM, counted and resuspended in buffer containing 20 mM HEPES (pH 7.0), 137 mM NaCl, 5 mM KCl, 0.7 mM  $\text{Na}_2\text{HPO}_4$ , 6 mM dextrose, and 0.1 mM  $\beta$ -mercaptoethanol. Just prior to electroporation, the oligonucleotides and lambda Beta protein are added to  $10^7$  ES cells in each 1 ml-cuvette. Cells and DNA are exposed to two sequential 625 V/cm pulses at room temperature, allowed to remain  
25 in the buffer for 10 minutes, then plated in non-selective media onto feeder cells.

Following two days of non-selective growth, the cells are trypsinized and replated onto G418 (250  $\mu$ g/ml) media. The positive-selection is applied for three days. Because of the high efficiency of lambda Beta-mediated recombination, the need for further selection (for example, negative selection by introducing a thymidine kinase

gene and selecting with ganciclovir) can be obviated. Appropriately transformed, G418-resistant cells are grown in non-selective media for 2-5 days prior to injection into blastocysts according to the method of Bradley in Teratocarcinomas and embryonic stem cells, a practical approach, edited by E. J. Robertson, IRL Press, Oxford (1987), p. 125.

Blastocysts containing the targeted ES cells are implanted into pseudo-pregnant females and allowed to develop to term. Chimeric offspring are identified by coat-color markers and those males showing chimerism were selected for breeding offspring. Those offspring which carry the mutant allele can be identified by coat color, and the presence of the mutant allele reaffirmed by DNA analysis by tail-blot, DNA analysis.

Thus, the method markedly simplifies the construction of gene knockouts and gene modifications in ES cells. In addition to its possible relevance to plant, animal and human gene therapy, the method will simplify the construction of transgenic animals harboring either gene knockouts or gene modifications.

As described for stem cells and in Example 14, a broad range of ssDNA and ssDNA binding polypeptide concentrations will be effective. For example, both ssDNA and ssDNA binding protein may be present in amounts ranging from 0.001  $\mu$ M to 100 mM; or from 0.1  $\mu$ M to 1  $\mu$ M; or from 1  $\mu$ M to 100  $\mu$ M. Oligonucleotide length can be varied in accordance with parameters presented in Example 12. There is no particular upper limit on oligonucleotide length. In addition, two or more oligonucleotides can be included which have complementary 5' ends (for example with 10, 20, 30, 40 bp complementary 5' ends), thereby creating 3' overhangs which are effective substrates for ssDNA binding polypeptides such as lambda Beta. In addition, RecT, P22 Erf, Rad52, and other double strand break repair ssDNA binding polypeptides may be substituted for lambda Beta, in the same ranges described for lambda Beta. Those skilled in the art will recognize that culture and electroporation conditions are readily variable without materially reducing homologous recombination. Moreover, nucleic acid may be introduced into the cell by any suitable macromolecular transfer method.



**EXAMPLE 16****Homologous Recombination in Plants**

The methods disclosed herein are also applicable to the manipulation of plant cells and ultimately the genome of the entire plant. A wide variety of transgenic plants have been reported, including herbaceous dicots, woody dicots and monocots. For a summary, see Gasser, et al. (1989), *Science*, 244, 1293-1299. A number of different gene transfer techniques have been developed for producing such transgenic plants and transformed plant cells. One technique used *Agrobacterium tumefaciens* as a gene transfer system. Rogers, et al. (1986), *Methods Enzymol.*, 118, 627-640. A closely related transformation utilizes the bacterium *Agrobacterium rhizogenes*. In each of these systems a Ti or Ri plant transformation vector can be constructed containing border regions which define the DNA sequence to be inserted into the plant genome. These systems previously have been used to randomly integrate exogenous DNA to plant genomes.

Preferably, DNA designed for homologous recombination with a target DNA sequence in plants are combined with lambda Beta protein or other ssDNA protein and directly transferred to plant protoplasts by way of methods analogous to that previously used to introduce transgenes into protoplasts. Concentration of the DNA and ssDNA binding proteins are as described in Example 15. See, e.g. Paszkowski, et al. (1984), *EMBO J.*, 3, 2717-2722; Hain, et al. (1985), *Mol. Gen. Genet.*, 199, 161-168; Shillito, et al. (1985), *Bio/Technology*, 3, 1099-1103; and Negrutiu, et al. (1987), *Plant Mol. Bio.*, 8, 363-373. Alternatively, the PNS vector is contained within a liposome which may be fused to a plant protoplast (see, e.g. Deshayes, et al. (1985), *EMBO J.*, 4, 2731-2738) or is directly inserted to plant protoplast by way of intranuclear microinjection (see, e.g. Crossway, et al. (1986), *Mol. Gen. Genet.*, 202, 179-185, and Reich, et al. (1986), *Bio/Technology*, 4, 1001-1004). Microinjection is the preferred method for transfecting protoplasts. The DNA and ssDNA binding proteins may also be microinjected into meristematic inflorescences. De la Pena et al. (1987), *Nature*, 325,

274-276. Finally, tissue explants can be transfected by way of a high velocity microprojectile coated with the DNA and ssDNA binding proteins analogous to the methods used for insertion of transgenes. See, e.g. Vasil (1988), *Bio/Technology*, 6, 397; Klein, et al. (1987), *Nature*, 327, 70; Klein, et al. (1988), *Proc. Natl. Acad. Sci. U.S.A.*, 85, 8502; McCabe, et al. (1988), *Bio/Technology*, 6, 923; and Klein, et al., *Genetic Engineering*, Vol 11, J. K. Setlow editor (Academic Press, N.Y., 1989). Such transformed explants can be used to regenerate for example various serial crops. Vasil (1988), *Bio/Technology*, 6, 397.

Once the DNA and ssDNA binding protein have been inserted into the plant cell by any of the foregoing methods, homologous recombination targets the oligonucleotide to the appropriate site in the plant genome. As in previous examples, the oligonucleotide may be a series of overlapping ssDNAs with 5' or 3' overhangs. Depending upon the methodology used to transfect, selection is performed on tissue cultures of the transformed protoplast or plant cell. In some instances, cells amenable to tissue culture may be excised from a transformed plant either from the F0 or a subsequent generation.

The amino acid composition of various storage proteins in wheat and corn, for example, which are known to be deficient in lysine and tryptophan may also be modified. PNS vectors can be readily designed to alter specific codons within such storage proteins to encode lysine and/or tryptophan thereby increasing the nutritional value of such crops. For example, the zein protein in corn (Pederson et al. (1982), *Cell*, 29, 1015) may be modified to have a higher content of lysine and tryptophan by the vectors and methods of the invention.

**EXAMPLE 17****Creation Of Improved Bacterial Host Strains for Lambda-Mediated  
Recombination in BACs****5    *Transfer of lambda recombination genes to DH10B cells.***

To facilitate the use of lambda-mediated recombination with BACs, an improved phage-mediated recombination system has been created for efficient recombination using BACs. The DH10B strain unlike most other strains of *E. coli* is efficiently transformed with BAC DNA and contains many of the BAC genomic  
10    libraries; it was judged to be a good host strain for subsequent modification.

Because DH10B is *recA* defective, standard genetic crosses cannot be used to place the defective lambda prophage used for lambda-mediated recombination into the DH10B strain of *E. coli*. To circumvent this problem, DH10B was first converted to *recA*<sup>+</sup>, then the lambda recombination genes were crossed in and the strain was again  
15    made *recA*<sup>-</sup> but now carrying the lambda genes.

To make DH10B *recA*<sup>+</sup>, a lambda transducing phage carrying the *recA*<sup>+</sup> gene was used to lysogenize DH10B creating the derivative DY303. In strain DY303 used for  $\lambda$  mediated recombination, the *tet* gene conferring tetracycline resistance was inserted by homologous recombination where the *cro-bioA* deletion exists creating  
20    strain DY363. A P1 lysate made on DY363 was used to infect DY303 and by standard bacterial genetics the *tet* gene in the *cro-bioA* deletion was crossed into DY303. This deletes a large segment of the  $\lambda$  DNA of the lysogen including the *recA*<sup>+</sup> gene. This new derivative of DH10B and DY303 is named DY380, is *RecA*<sup>-</sup> and carries the *tet* selectable marker substituted for the *cro-bioA* segment.. It was observed that DY380  
25    cells were transformed with BAC DNA at efficiencies of  $10^{-6}$  to  $10^{-4}$ .

*Creation of DY380 derivatives containing arabinose-inducible cre or flpe genes*

BAC targeting often makes use of a selectable marker to introduce the targeting cassette into the targeted locus. The selectable marker can, however, interfere with the subsequent function of the targeted locus. To circumvent this problem, the inventors  
 5 noted that a selectable marker flanked with either *FRT* or *loxP* sites can be removed by either Flp or Cre recombinases, respectively. Thus, the inventors have created two new strains, EL250 and EL350, by ultimately replacing the *tet* gene in DY380 with *araC* and placing the *flpe* and *cre* genes under an arabinose-inducible promoter. The genotypes of DY380, EL250, and EL350 are shown schematically in FIG. 6. Although the  
 10 arabinose-inducible promoter *pBAD* was used in this example, essentially any inducible promoter may be used to activate *flpe* and *cre* expression.

In DY380's prophage, *tet* is located between *ci857* and *bioA*. In EL250's prophage, *flpe* replaces *tet* (*flpe* is a genetically engineered *flp* that has a higher recombination efficiency than the original *flp* gene; Buchholz *et al.*, Nature  
 15 Biotechnology 16: 657-662, 1998). Thus, as illustrated in FIG. 6, both EL250 and EL350 have heat-inducible homologous recombination (the  $\lambda$  *red* genes) and arabinose-inducible site-specific recombination (*flpe* or *cre*) functions. This dual regulation allows both selective targeting by recombination as well as the subsequent removal of the selection marker from the targeted locus by site-specific recombination.

20 *Improved approach for introducing defective lambda prophages into bacteria: mini lambda circles*

A method has been developed for introducing the  $\lambda$ -mediated recombination system directly into nearly any *E. coli* strain including *recA* defective DH10B derivatives. These derivatives may carry BACs, PACs, or other vectors.

25 The DY330 strain carries deletion of prophage genes from *cro* through *bioA*. This deleted segment of  $\lambda$  and *bioA* were replaced to create a derivative that contains a fully normal  $\lambda$  *ci857* single-copy lysogen. Lysogens of this type can be induced at 42° to express  $\lambda$  functions including the Red recombination functions. Because the  $\lambda$

carries all of the replication and lytic genes, induction for longer than 6 minutes causes death of cells carrying the lysogen. However, by inducing for less than 6 minutes, for example 4 minutes, recombination functions are only partially activated, but cells survive when returned to grow at 32°C. Using, for example, a 4-minute time of induction,  $\lambda$ -mediated recombinants can be generated between linear, electroporated DNA and the chromosome including the DNA of the prophage. Thus, phage lambda itself can be used to lysogenize and generate recombinants in BAC strains. However, recombination efficiency would be low because of the short induction time.

10 PCR cassettes containing 5 genes for different drug resistance markers were amplified [*cat*, *kan*, *amp*, *tet*, *spec* (*strep*)] with flanking homologies so as to replace prophage genes from *cro* through *ea59* with the respective drug markers selecting with that drug for resistant recombinants at 32°C. A contiguous prophage DNA segment from base position 38,044 of the  $\lambda$  map in *cro* to base position 25,737 of the  $\lambda$  map in *sib* are replaced by the drug cassettes. (see Court and Oppenheim). This deletion eliminates all replication and lysis genes of the prophage creating a defective prophage similar to that of the original DY330. The difference is that this prophage has both attachment sites *attL* and *attR* at the termini of the prophage whereas DY330 has *attR* through *bioA* deleted as part of the *cro-bioA* deletion.

20 This set of strains (with respective drug cassettes) can be induced for longer times than the complete lambda without killing the cells thereby providing maximal homologous recombination activity just as with DY330. The  $p_L$  operon of these prophages include the *int* and *xis* genes. Induction activates their expression and because both *attL* and *attR* are present causes site specific excision of the prophage as a DNA circle carrying its associated drug marker. Cells undergoing induction for 15-min may lose the original prophage. This happens in about 50% of the cells. The other 50% still have the prophage. The 50% with the prophage are likely to occur by reintegration of the circular DNA at the vacated *attB* site through Int-mediated site specific recombination.

The defective prophage DNA can be isolated and purified from these lysogens, if after a 15 minute induction, cells are lysed and DNA is isolated by plasmid purification protocols, ie by Qiagen columns. The circular phage DNA with its drug markers can be purified. This DNA cannot replicate upon retransfection into *E. coli* strains but it can express its pL operon and Int function to allow integration of the circular DNA by site specific recombination between *attP* in the circular DNA and *attB* in the bacterial chromosome. Only Int and the host IHF functions are required for sitespecific recombination. Such integrated DNAs are stable, are immune and can be selected by the drug marker each carries.

10 Because RecA is not required for site specific Int-mediated recombination, DH10B derivatives can be used for transformation and for integration of the circular defective phage DNA selecting for its appropriate drug marker.

15 The defective mini-prophage can also be induced as part of a di-lysogen in which a complete  $\lambda$  cI857 phage is also present. The phage lysate created by this 90 min induction at 42 degrees in L-Broth generates normal lambda phage particles as well as particles that contain the defective mini-prophage DNA (In  $\lambda$  terminology *docL* particles). Infection of these lysates into cells (e.g. DH10B) allows DNA injection of the mini-prophage DNA, site specific recombination, and selection for the drug marker  
20 carried on that DNA.

### EXAMPLE 18

#### An Improved Strategy and Improved Reagents for BAC Engineering

To test the prophage system of Example 17 in BAC engineering, the efficiency of BAC recombination in EL250 cells was investigated. In the experiments described in this  
25 example, a selectable cassette was targeted to a mouse neuron-specific locus in a 250 kb BAC. The BAC was then further modified to enhance its usefulness in subsequent mouse

genetic studies. These experiments validated an improved strategy and provided improved reagents for BAC engineering using the lambda recombination system.

The *Eno2* gene is located in the middle of 284H12, a fully sequenced BAC (obtained from Research Genetics; Ansari-Lari *et al.*, Genome Research 8: 29-40, 1998).  
5 The *Eno2* gene was targeted because it is neural-specific and expressed in most mature neurons (Marangos and Schmechel, Annual Review of Neuroscience 10: 269-295, 1987). By knocking out *eno2* and replacing it with a *cre*-containing cassette, a BAC transgenic line that expresses Cre in all mature neurons was created (described in Example 20). This BAC transgenic line is useful for subsequent conditional knockout studies. The inventors  
10 used a BAC approach in part because BACs are large enough to contain all the important regulatory sequences required for proper regulation of gene expression.

The following describes the construction of the BAC transgenic line with neuronal-specific Cre expression.

*Generation of the targeting cassette and BAC-containing EL250 cells*

15 The *IRES-eGFPcre-FRT-kan-FRT* targeting cassette was PCR amplified from pICGN21, which was constructed by subcloning a 1.9 kbp *HindIII/AccI*-digested and filled-in *FRT-kan-FRT* fragment from pFRTneo into the *NotI/BclI*-digested and filled-in cloning site of pIRESeGC. The *IRES-eGFPcre-FRT-kan-FRT* cassette was amplified using chimeric 63 nt primers. The 3' 21 nt of each primer was homologous to the  
20 targeting cassette used for amplification while the 5' 42 nt was homologous to the last exon of *Eno2* where the cassette was to be targeted by recombination. The primers were designed to precisely target the cassette downstream of the *Eno2* stop codon and upstream of its polyA site.

The *Eno2*-containing 284H12 BAC was electroporated into EL250 cells and six  
25 chloramphenicol resistant ( $\text{Cm}^R$ ) colonies selected. Digestion of BAC DNA from six  $\text{Cm}^R$  colonies with *EcoRI* or *HindIII* showed that one had an abnormal digestion pattern. However, in other BAC electroporation experiments involving the analysis of more than 76 additional colonies, no abnormal BACs were identified. These results indicate that

BAC rearrangements during electroporation are rare. Subsequent experiments were carried out with Cm<sup>R</sup>-resistant EL250 colonies harboring BACs having proper *EcoRI* and/or *HindIII* digestion patterns.

*Generating and isolating a BAC with a disrupted eno2 locus*

5        Next, the 284H12 BAC was modified to disrupt the *eno2* locus with the *IRES-eGFPcre-FRT-kan-FRT* targeting cassette. The methods used in these experiments were similar to those described extensively herein, and in Yu *et al.* (Proc. Natl. Acad. Sci. U.S.A. 97: 5978-5983, 2000). The approach is illustrated schematically in FIG. 7.

10        EL250 cells carrying the 284H12 BAC were shifted to 42°C for 15 min. to induce lambda Exo, Beta, and Gam expression. The cells were then electroporated with 300 ng of the amplified *IRES-eGFPcre-FRT-kan-FRT* cassette, and kanamycin-resistant (Km<sup>R</sup>) colonies were selected. A kanamycin-resistant phenotype indicated that the targeting cassette was successfully integrated into the 284H12 BAC (illustrated as "Targeting," FIG. 7, middle). Approximately 5200 Km<sup>R</sup> colonies were obtained from 10<sup>8</sup> electroporated  
15        cells for a targeting efficiency of about 10<sup>-5</sup>. No colonies were obtained from control cells that were not heat-induced. Thus, lambda recombinase expression was required for efficient recombination.

20        Twenty-four kanamycin resistant colonies were analyzed with whole-cell PCR using primers that flanked the targeted locus. The PCR results indicated that all were correctly targeted. Sequencing of the targeted region from six colonies, however, showed that three carried point mutations. To determine whether these point mutations were introduced during PCR amplification or during homologous recombination, the targeting was repeated. This time, however, the PCR-amplified *IRES-eGFPcre-FRT-kan-FRT* cassette was subcloned into the *SmaI* site of pBluescript by blunt-end ligation before  
25        targeting, and plasmids carrying wild type amplified cassettes identified by DNA sequencing. These cassettes were then released from the plasmid by *BamHI* digestion and used for targeting. Using this two-step method, all twelve targeted BACs that were subsequently sequenced contained wild type *IRES-eGFPcre-FRT-kan-FRT* cassettes.



These results indicate that the point mutations were introduced by the primers used or during PCR amplification of the targeting cassette rather than during targeting.

*Removing the kanamycin-resistance marker*

Next, the *kan* selectable marker was removed from the BAC to prevent it from possibly interfering with Cre expression. This process was initiated by arabinose treatment, which induces EL250 cells to express the Flpe recombinase. The process is illustrated in Fig. 7, bottom line ("Flip-out of *kan*").

Overnight cultures from single *Km<sup>R</sup>* colonies were diluted 50-fold in LB medium and grown till  $OD_{600} = 0.5$ . Flpe expression from the EL250 cells was then induced by incubating the cultures with 0.1% L-arabinose for 1 hr. The bacterial cells were subsequently diluted 10-fold in LB medium, grown for an additional hour, and spread on chloramphenicol plates (12.5 ug/ml). The next day, 100 *Cm<sup>R</sup>* colonies were picked and replated on *kan* plates (25 ug/ml) to test for loss of kanamycin resistance. Chloramphenicol resistance indicates that the cell retained the BAC, whereas kanamycin sensitivity indicates that *kan* has been successfully removed from the BAC. All colonies were *Km<sup>S</sup>* and contained a single *FRT* site at the targeted locus.

Without being bound by theory, it is likely that the surprisingly high recombination efficiency reflects the tight control of Flpe expression afforded by the single copy *P<sub>BAD</sub>* promoter and *flpe* gene, and the fact that the *FRT* sites are located in *cis* rather than in *trans* to each other.

*Removing an undesirable loxP site in the BAC vector backbone*

A *loxP* site contained in the BAC vector backbone (pBeloBAC11; Shizuya *et al.*, Proc. Natl. Acad. Sci. U.S.A. 89: 8794-8797, 1992) was removed by a final round of gene targeting.

To facilitate the removal of this undesirable *loxP* site, a new plasmid, pTamp, was constructed that contains an *amp* gene flanked by 920 bp of pBeloBAC11 sequence located 5' of the *loxP* site and 370 bp of pBeloBAC11 sequence located 3' of the *loxP* site.

This *amp* insert can be released from pTamp by *Bam*HI digestion and used to replace the *loxP* site in the BAC transgene by gene targeting. This targeting reaction is very efficient due to the large amount of homology between the *amp* cassette and the pBeloBAC11 vector (56,200 colonies per  $10^8$  electroporated cells).

- 5        Upon removal of the undesirable *loxP* site, the modified 284H12 BAC was used in the transgenic mouse studies described in Example 20.

### EXAMPLE 19

#### Subcloning by GAP Repair

- 10        This  $\lambda$ -mediated recombination system can also be used to subclone fragments from BACs without the use of restriction enzymes or DNA ligases. This form of subcloning relies on gap repair to recombine the free ends of a linear plasmid vector with homologous sequences carried on the BAC (FIG. 8). The method is readily adaptable to other forms of intramolecular and extrachromosomal DNA, such as plasmids, yeast artificial chromosomes, P1 artificial chromosomes, and cosmids. This novel method combines
- 15        lambda mediated recombination with gap repair to enable recombination of very large DNA segments onto an extrachromosomal vector.

- 20        The linear plasmid vector with an *amp* selectable marker and an origin of replication carries the recombinogenic ends (FIG 8B). The vector is generated by PCR amplification using two chimeric primers. The 5' 45-52 nt of each primer is homologous to the two ends of the BAC sequence to be subcloned while the 3' 20 nt is homologous to plasmid DNA. Recombination generates a circular plasmid in which the DNA insert was retrieved from the BAC DNA via gap repair. Circular plasmids are selected by their  $\text{Amp}^R$ .

- 25        To determine the maximum sized fragment that can be subcloned from BACs using this method, several different pairs of primers were generated in which the homology segments were located 25 kb, 48 kb, 60 kb, or 80 kb apart in the *Eno2* BAC DNA (Fig. 8A). Rare cutter *Not*I and *Sal*I restriction sites were also incorporated into these primers

so that the subcloned fragments could be released from the recombinant clones intact. Using pBluescript as the cloning vector, it was possible to subclone the 25 kb fragment. However, attempts to subclone larger fragments were unsuccessful. As a possible explanation for this result, it was hypothesized that subclones containing larger fragments  
5 on a high copy vector such as pBluescript were toxic to the cell.

To determine if the hypothesis was correct, a lower copy number vector (pBR322, with its copy number control element intact) was used as the cloning vector. Fragments as large as 80 kb could be subcloned with a pBR322 vector. Not all subclones obtained by gap repair had the correct inserts (as determined by restriction enzyme pattern analysis).  
10 Some subclones lacked inserts while others contained inserts with aberrant restriction patterns. In order to confirm that the correct insert has been subcloned, when using subcloning by gap repair, a method of screening subclones can be used to assure that the selected subcloned contains the desired insert. Such methods include restriction mapping, sequencing, PCR analysis, Southern analysis, etc., and other methods well known to those  
15 of skill in the art.

The ability to subclone large fragments of genomic DNA by gap repair should facilitate many studies in genome research that were difficult or impossible to perform previously. For example, Gap repair for cloning on to vectors can be used with many different vectors used for protein expression in bacteria, plants and animal cells;  
20 mutagenesis; cloning; transcription; etc. Targeting vectors or transgenic constructs can be subcloned with ease, and virtually any region of the engineered BAC can now be included in the desired subclone.

Lambda mediated recombination combined with gap repair makes it possible to subclone fragments from complex mixtures without first purifying the DNA to be  
25 subcloned. This greatly facilitates the subcloning process and allows for high throughput subcloning of tens of thousands of genes or DNA molecules into many different vector backbones. This will greatly facilitate studies designed to determine the function of genes uncovered in large scale sequencing projects. For example, cDNA clones for genes of unknown function can be subcloned into many different expression vectors and the

function of these genes studies in cell-based assays in vitro or in the whole animal. This type of subcloning does not rely on PCR amplification, which can introduce unwanted mutations into the subcloned sequences.

Subcloning by gap repair also facilitates the identification of locus control regions or other regulatory elements that may be located at some distance from the gene. Many such potential elements are presently being identified by techniques such as comparative genome sequencing. Examples include pathogenicity islands, replicative origins and segregation elements. The ability to modify precisely these regulatory sequences on BACs, combined with the ability to include or exclude them during the subcloning process, will make it possible to dissect the function of these sequences in the whole animal or in vitro at a level not previously possible.

## EXAMPLE 20

### Production of Transgenic Mice Using BACs

Examples 17 and 18 describe the construction of a modified BAC believed to contain all of the regulatory sequences needed for neural-specific Cre expression in transgenic mice. To investigate this hypothesis, the modified BAC described in Example 18 was injected into (C3H/HeN-Mtv<sup>-</sup> X C57BL/6Ncr) F<sub>2</sub> zygotes. A BAC transgenic line carrying approximately two copies of the transgene was then established.

In addition to the BAC transgenic line, two transgenic lines carrying 25-kbp subclones of the BAC were also established. The 25-kbp subclones contains the entire modified *Eno2* coding region as well as 10 kbp of 5' and 5 kbp of 3' flanking sequences, respectively. One transgenic line, 25kbp-1 carries approximately four copies of the transgene, while the second, 25kbp-2 carries approximately five copies of the transgene. Thus, Cre expression in the BAC transgenic line could be compared to Cre expression in the transgenic lines carrying the subclone.

The transgenic mice were crossed to ROSA26 reporter mice, which contain a lacZ reporter that can be activated by Cre recombinase (Soriano, Nature Genetics 221:70-71,

1999). Double heterozygotes were subsequently analyzed by X-gal staining at 4 weeks of age.

Several different tissues were examined for X-gal expression including the brain, spinal cord, eye, lung, heart, intestine, muscle, liver, spleen, and kidney. Blue stained cells were found only in neural tissue in the three transgenic lines, indicating that both the BAC and the 25-kbp subclone contain the regulatory elements needed for neural-specific expression. The pattern of Cre activity was, however, different in the three lines. Vibratome sections of the brain from the BAC transgenic mice showed blue-stained cells throughout the gray matter but not in the white matter, indicative of Cre activity in most neurons but not in glial cells. In contrast, X-gal staining in the 25kbp-1 and 25kbp-2 transgenic mice was present in only a subset of neurons and expression was variable between the two different lines.

Higher power magnification of the cerebellum of the BAC transgenic mice showed that Cre was expressed in virtually all neuronal cells. This included Purkinje cells in the Purkinje cell layer, granule and Golgi cells in the granular layer, basket cells and stellate cells in the molecular layer and neurons of the deep cerebellar nuclei. In contrast, in the 25kbp-1 line, Cre was expressed in only a subset of Golgi cells in addition to a few cells in the granule and Purkinje cell layers. Glial cells of white matter also expressed Cre indicative of leaky expression. In the 25kbp-2 line, Cre expression was limited to the gray matter and included a variety of neuronal cell types, including most basket cells, stellate cells, Purkinje cells and neurons of the deep cerebellar nuclei. In contrast, few granule cells and Golgi cells in the granule layer expressed Cre.

Higher power magnification of the hippocampus and cortex showed similar results. In the hippocampus of BAC transgenic mice, virtually all neurons in the cornu Ammonis (CA) region and the dentate gyrus (DG) expressed Cre. The same was true in the cortex, where all six layers of the cortex that contained neurons (layers II-VI) expressed Cre. In contrast, the hippocampus of 25kbp-1 transgenic mice showed reduced Cre expression in the DG (Fig. 4E) and layers II and III of cortex. The 25kbp-2 transgenic mice showed even lower levels of Cre expression in the DG. The CA1 and CA2 regions of the CA also

failed to express Cre. Cre expression was also greatly reduced in the cortex, with layers II and III showing most the reduction.

Cre activity in the spinal cord, dorsal root ganglion (DRG) and retina of the transgenic mice was also examined in order to determine whether Cre was expressed in  
5 mature neurons within the peripheral nervous system. Similar to what was observed for the central nervous system, Cre was expressed in most mature peripheral neurons in the BAC transgenic mice while fewer peripheral neurons expressed Cre in the two 25kbp transgenic lines (data not shown).

To determine whether Cre was expressed in all Eno2 protein-positive neurons, a  
10 section from the brain of a BAC transgenic animal was immunostained with an anti-Eno2 antibody followed by X-gal staining for Cre activity. Virtually all Eno2-positive neurons were active for Cre. Thus, Cre expression in BAC transgenic animals correlated tightly with native mouse Eno2 promoter-enhancer activity.

The present application, particularly Examples 17-20, describes a highly efficient  
15 recombination system for manipulating BAC DNA in *E. coli*. The recombination system uses a defective  $\lambda$  prophage to supply functions that protect and recombine the electroporated linear DNA targeting cassette with the BAC sequence. Because the recombination functions are expressed from a defective prophage rather than a plasmid, the recombination functions are not lost during cell growth as often happens with plasmid-  
20 based systems. Another advantage of this prophage system is that the  $\lambda$  *gam* and *red* recombination genes are under the control of the temperature sensitive  $\lambda$  repressor that provides a much tighter control of *gam* and *red* expression than can be obtained on plasmids. This tight regulation, combined with the strong  $\lambda$  *PL* promoter, which drives *gam* and *red* expression to very high levels, makes it possible to achieve recombination  
25 frequencies that are surprisingly efficient (at least 50-100 fold higher than those obtained with plasmid-based systems; Narayanan, et al., Gene Therapy 6: 446: 442-447, 1999; Muyers et al., Nucleic Acids Research 27: 1555-1557, 1999). The tight control prevents expression of any recombination functions except for the 15min temperature induction.

The ability to precisely manipulate large fragments of genomic DNA, independent of the location of appropriate restriction enzyme sites, has many applications for functional genomics, both in the mouse and in other organisms. As shown herein, Cre can be introduced into the coding regions of genes carried on BACs facilitating the generation of

5 Cre-expressing transgenic lines for use in conditional knockout studies or for use in conditional gene expression studies. Genes can also be epitope tagged and microinjected into the germline of mice carrying a mutation in the gene. If the epitope tagged transgene rescues the mutant phenotype, the epitope tagged protein is functional and the epitope tag can serve as a marker for expression of the gene. Likewise, a gene carried on a BAC can

10 be replaced with another gene and the function of the "knock-in" mutation assayed in transgenic mice.

This recombination system also facilitates the generation of complicated conditional targeting vectors. While the generation of such vectors often used to take several months it can now be done in a only few weeks time. The ability to reversibly

15 express Cre or Flpe recombinases in *E. coli* speeds this process even further. Moreover, as demonstrated in Example 18, a selectable marker flanked with *loxP* or *FRT* sites can be now be introduced into an intron of a gene and then removed by transient Cre or Flpe expression leaving behind a solo *loxP* or *FRT* site in the intron.

## EXAMPLE 21

### 20 BAC Recombination without Drug Selection

The high efficiency of recombination described in Example 18 and elsewhere in these examples suggested that it might be possible to do targeting without drug selection. Direct targeting without drug selection would offer a number of significant advantages. In particular, it would facilitate genomic experiments in which the presence of a selectable

25 marker, or even a *FRT* or *loxP*, scar might be undesirable.

To demonstrate that targeting can be achieved without drug selection, a 24 bp FLAG tag was targeted to the 5' end of the SRY-box containing gene 4 (*Sox4*) gene

carried on a 125 kb BAC. For these experiments, a 114 bp targeting cassette was generated in which two 45-bp arms homologous to the *Sox4* gene flanked the 24-bp FLAG sequence. This DNA fragment was created by synthesizing two 79-bp oligonucleotides that overlapped at their 3' ends by 44 bp. These overlaps were annealed and filled in by  
5 Taq polymerase.

Expression of lambda recombinase genes from the defective prophage was heat-induced in DY380 cells carrying the *Sox4* BAC. Then, the FLAG-tagged cassette was introduced into the cells by electroporation. The cells were then spread on LB plates to a density of ~2,000 cells per plate. Colonies containing the FLAG tag were subsequently  
10 identified by colony hybridization using a 30-bp FLAG-specific oligonucleotide probe (24 bp FLAG tag and 3 bp on each side that was homologous to the *Sox4* targeted site).

Among 3,800 colonies screened from uninduced cells, no FLAG-positive colonies were identified. In contrast, seven FLAG-positive colonies were identified in 4,210 colonies obtained from induced cells for an overall targeting frequency of  $1.7 \times 10^{-3}$ . PCR  
15 amplification and direct sequencing showed that each of the seven FLAG-positive colonies was correctly targeted.

As unequivocally demonstrated in this example, the surprisingly high recombination efficiency offered by this recombination system makes it possible to manipulate BAC or other DNA without drug selection. Point mutations, deletions, or  
20 insertions can now be engineered into any gene on a BAC in the absence of a confounding linked drug selection marker or a *loxP* or *FRT* site. In cases where the gene is mutated in human disease, the exact disease-causing mutations can be engineered on the BAC and the effect of these mutations analyzed in transgenic mice.

## EXAMPLE 22

### Materials and Methods Used in Examples 17-21

25 *Bacterial strains.* All of the strains used except DH10B were maintained at 32°C because of the temperature inducible prophage. DY303 was constructed by infecting



DH10B cells (Gibco) with a  $\lambda$  phage carrying *recA* ( $\lambda$ cl857 *recA*<sup>+</sup>) (a gift from F. W. Stahl) and lysogens were selected. Strain EL11 was constructed by replacing the *tet* gene of DY380 with a cassette containing the *cat* and *sacB* genes by selecting Cm<sup>R</sup>. EL11 cells are Tet<sup>S</sup>, Cm<sup>R</sup> and sensitive to 2% sucrose. Strain EL250 was constructed by replacing the *cat-sacB* cassette of EL11 cells with *araC* and the arabinose promoter-driven *flpe* recombinase gene (*P*<sub>BAD</sub>*flpe*) selecting in the presence of sucrose. EL250 cells are resistant to 2% sucrose. Strain EL350 was constructed in a similar manner except for *cre* instead of *flpe*.

*Construction of plasmids.* The *IRES-eGFPcre-FRT-kan-FRT* targeting cassette was PCR amplified from pICGN21, which was constructed by subcloning a 1.9 kbp *HindIII/AccI*-digested and filled-in *FRT-kan-FRT* fragment from pFRTneo into the *NotI/BclI*-digested and filled-in cloning site of pIRESeGC. The FRTneo was constructed by amplifying the *kan* gene along with the Beta lactamase promoter from pEGFP-C1 (Clontech) with primers

5'CTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTTCGTCAGGTGGC  
 ACTTTCGGG (SEQ ID NO:13) and 5'CTCAGAAGAACTCGTCAAGAAGG (SEQ ID  
 NO:14). The amplified fragment was then targeted between the *FRT* sites in pNeo $\beta$ -gal (Stratagene). The pIRESeGC was generated by inserting the 2 kbp *NheI/MluI*-digested and filled-in eGFPcre fragment from pEGC into the 3.5 kbp *BamHI*-digested and filled-in cloning site of pNTRLacZPGKneoloxP (Arango *et al.*, 1999). The pEGC was generated by subcloning a 1.05 kbp *EcoRI/KpnI* PCR fragment containing the *cre* gene from pGKmncre (a gift from P. Soriano) into the *EcoRI/KpnI* site of pEGFP-C1. This PCR fragment was generated by amplifying the *cre* gene from pGKmncre with primers

5'GTAGGTACCTCGAGAATCGCCATCTTCCAGCAGGC (SEQ ID NO:15) and  
 5'TCGAATTTTCTGCATCCAATTTACTGACCGTACACC (SEQ ID NO:16), which  
 contain *EcoRI* and *KpnI* cleavage sites, respectively, at their 5' ends.

To construct the pTamp vector, the *amp*-targeted pBeloBAC11 was first generated by replacing the *loxP* site in pBeloBAC11 (Shizuya *et al.*, 1992) with the PCR amplified *amp* gene from pEGFP (Clontech). The primers used for amplification are

5' *GCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGTCTTAGA*  
*CGTCAGGTGGCAC* (SEQ ID NO:17) and  
 5' *CATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGAACCTCACGT*  
*TAAGGGATTTTGGTC* (SEQ ID NO:18), which are homologous to the *amp* gene of  
 5 pEGFP (in plain) and to sequences flanking the *loxP* site in pBeloBAC11 (in italic). A  
 2.4-kbp PCR fragment amplified from *amp*-targeted pBeloBac11 with primers  
 5'GCAGGATCCAGTTTGCTCCTGGAGCGACA (SEQ ID NO:19) and  
 5'TGCAGGTCGACTCTAGAGGATC (SEQ ID NO:20) was then cloned into the  
*XhoI/XbaI* and filled-in site of pCS (Stratagene) to create the pTamp vector. The 2.4 kbp  
 10 *amp* cassette containing an *amp* gene along with 920 bp of 5', and 370 bp of 3',  
 pBel0BAC11 vector sequence flanking the *loxP* site can be released by *Bam*H1 digestion  
 and used directly to replace the *loxP* site in any pBeloBAC11-derived BACs with *amp*.

The pKO4 vector containing the *cat-sacB* targeting cassette is a derivative of  
 pKO3 (Link *et al.*, 1997) in which 605 bp had been deleted between *cat* and *sacB*.

15 The *araC*-P<sub>BAD</sub>/*flpe* targeting cassette was amplified from pBAD*flpe*, which was  
 constructed by subcloning a 1.4 kb *PstI/KpnI* fragment from pOGF*lpe* (Buchholz *et al.*,  
 1998) into pBAD/MycHis-A (Invitrogen). The *araC*-P<sub>BAD</sub>*cre* targeting cassette was  
 amplified from pBAD*cre*, which was constructed by introducing a 1.2 kb *HindIII/NcoI*  
 fragment from pGK*mncre* into pBAD/His-C (Invitrogene).

20 *Amplification primers for targeting or GAP repair cassette DNAs.* For all primers  
 listed below, nucleotides in italics are homologous to the targeted sequence, while those in  
 plain text are homologous to amplification cassettes. The Tet<sup>R</sup> cassette used for targeting  
*cro-bio* in DY330 was amplified from Tn10 with primers:

5'TGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTGTATGCTCTTGGGTTATCAA  
 25 GAGGG (SEQ ID NO:21) and  
 5'GGCGCTGCAAAAATTCTTTGTGGAACAGGGTGTCTGGATCACTCGACATCTTGGT  
 TACCG (SEQ ID NO:22). The *cat-sacB* cassette used for replacing the *tet* gene in DY363  
 was amplified from pKO4 with primers:

5'TGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTGTATGCTGTGACGGAAGAT

- CACTTCG (SEQ ID NO:23) and  
 5'GGCGCTGCAAAAATTCTTTGTGCGAACAGGGTGTCTGGATCCTGAGGTTCTTATGG  
 CTCTTG (SEQ ID NO:24). The *araC*-P<sub>BAD</sub>*flpe* and *araC*-P<sub>BAD</sub>*cre* cassettes used for  
 replacing the *cat-sacB* in EL11 were amplified from pBAD*flpe* and pBAD*cre* with  
 5 primers:  
 5'TGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTGTATGAAGCGGCATGCATA  
 ATGTGC (SEQ ID NO:25) and  
 5'GGCGCTGCAAAAATTCTTTGTGCGAACAGGGTGTCTGGATCCTGTGTCCTACTCAG  
 GAGAGCGTTC (SEQ ID NO:26). The *IRES-eGFPcre-FRT-kan-FRT* cassette used for  
 10 targeting the *Eno2* locus was amplified from pIGCN21 with primers:  
 5'CGCTTCGCGGGACATAATTTCCGAAATCCCAGTGTGCTGTGAGCCAAGCTATCGA  
 ATTCCGCC (SEQ ID NO:27) and  
 5'GAGGCTCCAGGAGAATGAGATGTTCCCGCGTTCAGGCAAGCGCTATTCCAGAAG  
 TAGTGAGGA (SEQ ID NO:28). The oligonucleotides used to target the flag cassette  
 15 into the 5' end of the *Sox4* gene were annealed and polymerase-extended using primers:  
 5'GCGAGCGTGTGAGCGCGCGTGGGCGCCCGGCAAGCCGGGGCCATGGATTACA  
 AGGATGACGACGATAAGGTACAACAGA (SEQ ID NO:29) and  
 5'GGCCAGCAGAGCCTCAGTGTTCTCCGCGTTGTTGGTCTGTTGTACCTTATCGTCG  
 TCATCCTTGTAATCCATGGCCCCC (SEQ ID NO:30).
- 20 The linear pBR322 derivative used to subclone the 25-kbp fragment from the  
 modified *Eno2* locus was amplified with primers:  
 5'CTCTCCATGCCTGTCTGGGTGAGGGTGGCCCAGGGGCGATGGCTATGAGAGAGG  
 TCGACTTCTTAGACGTCAGGTGGCAC (SEQ ID NO:31) (*Eno2*-C-L1) and  
 GCAATGCAGAGAAGCCTTGTA'CTGGGATGACAGAGACGGAGGGGAAGAGGAGGCG  
 25 GCCGCGATACGCGAGCGAACGTGA (SEQ ID NO:32) (*Eno2*-C-R1/2). The  
 amplification primers for the other experiments were: 48-kbp modified fragment,  
 5'GACTTCTATGACCTGTACGGAGGGGAGAAGTTTGCGACGTGACAGAGCTGGTCGT  
 CGACTTCTTAGACGTCAGGTGGCAC (SEQ ID NO:33) (*Eno2*-C-L2/3/4) and *Eno2*-C-  
 R1/2; 60-kbp modified fragment, *Eno2*-C-L2/3/4 and

5'GCCCCATACACGTAAATGTACATAGAATCACACAGCATCACTTCTATGGATGCGGC  
 GGCCGCGATACGCGAGCGAACGTGA (SEQ ID NO:34) (*Eno2*-C-R3); 80-kbp  
 modified fragment, *Eno2*-C-L2/3/4 and  
 5'CATCCAGTAGAACTTGGGAGTGAAGCTAGAGCCAAGGCCATCTAAGTGACAGGCG  
 5 GCGCGATACGCGAGCGAACGTGA (SEQ ID NO:35) (*Eno2*-C-R4). These primers  
 contained 5' regions homologous to the target sequence and 3' regions homologous to  
 pBR322. PCR products were purified using a Qiaex II gel extraction kit (Qiagen) and  
 digested with *DpnI* to remove contaminated template.

*Preparation of electrocompetent cells and generation of recombinants.* For BAC  
 10 modification, overnight cultures containing the BAC were grown from single colonies and  
 were diluted 50-fold in LB medium and grown to an  $OD_{600} = 0.5-0.7$ . 10-ml cultures were  
 then induced for Beta, Exo, and Gam expression by shifting the cells to 42°C for 15 min  
 followed by chilling on ice for 20 min. Cells then were centrifuged for 5 min at 5,500 g at  
 4°C and washed with 1.5 ml of ice-cold sterile water three times. Cells were then  
 15 resuspended in 50 µl of ice-cold sterile water and electroporated. For BAC transformation,  
 the induction step was omitted.

Cell transformation was performed by electroporation of 100-300 ng linear DNA  
 into 50 µl of ice-cold competent cells in cuvettes (0.1 cm) using a Bio-Rad gene pulser set  
 at 1.75 kV, 25 µF with pulse controller set at 200 ohms. 1 ml of LB medium was added  
 20 after electroporation. Cells were incubated at 32°C for 1.5 hr with shaking and spread on  
 appropriate selective or nonselective agar media.

*Production of transgenic mice.* Modified BAC and the p25-kbp subclone DNAs  
 were purified using cesium chloride gradients as described (Antoch *et al.*, 1997). The 25-  
 kbp subclone DNA was linearized by *NotI* digestion before microinjection. BAC DNA  
 25 (1 µg/ml) and 25-kbp subclone DNA (2 µg/ml) were microinjected into the pronucleus of  
 (C3H/HeN-Mtv<sup>+</sup> X C57BL/6Ncr) F<sub>2</sub> zygotes. Transgenic founders were subsequently  
 identified by Southern analysis using a *cre* probe or by PCR using primers  
 5'CTGCTGGAAGATGGCGATTCTCG (SEQ ID NO:36) and

5'AACAGCAGGAGCGGTGAGTC (SEQ ID NO:37) that flank the 3' insertional junction.

*Histochemical analysis of  $\beta$ -galactosidase expression.* Mice at 4 to 5 weeks of age were sacrificed in CO<sub>2</sub> and perfused with 4% paraformaldehyde in PBS (pH 7.3). The brains, spinal cords and eyes were removed and postfixed for 3 hr. Vibratome sections (20  $\mu$ m) of brains were mounted on slides and used directly for X-gal staining or for immunocytochemistry. For spinal cords and eyes, cryostat sections (20  $\mu$ m) were used that were made by cryoprotecting tissues in 30% sucrose in PBS overnight and embedding the tissues in freezing compound (OCT, Sakura). Before X-gal staining, samples on slides were postfixed with 0.25% glutaraldehyde in PBS and briefly washed with rinse solution (0.1 phosphate buffer pH7.3, 0.1% deoxycholic acid, 0.2% NP40 and 2mM MgCl<sub>2</sub>). X-gal staining was performed by incubating samples in staining buffer (2.5 mg/ml X-gal, 5mM potassium ferricyanide and 5 mM potassium ferrocyanide in staining buffer) for 2 hr at 37°C followed by counterstaining with 0.25% eosin (Fisher).

*Immunocytochemistry.* Immunostaining was carried out using the ABC Vectastain kit (Vector Labs) on 20  $\mu$ m vibratome sections. Sections were blocked with PBS (pH7.3 containing 0.2% Triton X-100, 1.5% bovine serum albumin and 5% normal goat serum) at room temperature for 2 hr and incubated with primary *Eno2* antibody, a poly clonal rabbit anti-*Eno2* antiserum (Chemicon) at 1:100 dilution in PBS solution. After incubation with a secondary biotinylated antibody and the ABC reagent, peroxidase was reacted with 0.05% diaminobenzidine tetrahydrochloride (DAB) and 0.003% hydrogen peroxide.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will

control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

## We Claim:

1. A method for performing homologous recombination of a nucleotide sequence of interest in a host cell, comprising  
introducing a nucleic acid sequence encoding Exo, Beta and Gam, or a functional variant or homologue thereof, operably linked to a repressible promoter into a population of host cells;  
introducing a polynucleotide comprising a nucleotide sequence homologous to the nucleotide sequence of interest into the host cells;  
activating the repressible promoter in the cells, thereby inducing the expression of Exo, Beta and Gam, or the functional variant or homologue thereof  
selecting a cell from the population of host cells in which homologous recombination has occurred between the nucleic acid of interest and a target nucleic acid sequence.
2. The method of claim 1, wherein at least one cell in 100,000 cells from the population of host cells contains DNA in which homologous recombination has occurred.
3. The method of claim 1, wherein at least one cell in 1,000 cells from the population of host cells contains DNA in which homologous recombination has occurred.
4. The method of claim 1, wherein at least one cell in 100 cells from the population of cells contains DNA in which homologous recombination has occurred.
5. The method of claim 3, wherein the nucleotide sequence of interest is intrachromosomal or extrachromosomal.

6. The method of claim 5, wherein the nucleotide sequence of interest is extrachromosomal and located on a bacterial artificial chromosome, a P1 artificial chromosome, plasmid, cosmid, or a yeast artificial chromosome.
7. The method of claim 1, wherein the homologous recombination alters a function of a gene in the cell.
8. The method of claim 1, wherein the nucleotide sequence homologous to the nucleotide sequence of interest is at least 30 base pairs in length.
9. A method for inducing homologous recombination in a cell comprising a target nucleic acid, the method comprising:
  - introducing into a cell a first nucleic acid of interest comprising a homologous sequence to the target nucleic acid of sufficient length to induce homologous recombination, wherein the cell comprises a nucleic acid encoding a single stranded DNA binding polypeptide that functions in double strand break repair homologous recombination, operably linked to a de-repressible promoter; and
  - inducing sufficient expression of the nucleic acid encoding the single stranded DNA binding polypeptide to promote homologous recombination of the first nucleic acid of interest with the target nucleic acid;
  - thereby inducing homologous recombination in the cell.
10. The method of claim 9, wherein the first nucleic acid of interest is a single stranded DNA and or a DNA comprising a 3' overhang.
11. The method of claim 10, wherein the homologous sequence is at least 70 nucleotides in length.
12. The method of claim 9, wherein the first nucleic acid of interest is a double stranded DNA comprising a homology arm.



13. The method of claim 12, wherein first nucleic acid of interest comprises a homology arm of at least 20 nucleotides in length.

14. The method of claim 9, wherein the single stranded DNA binding polypeptide is selected from the group consisting of: lambda Beta, RecT, P22 Erf, and Rad52.

15. The method of claim 9, wherein the single stranded DNA binding polypeptide is lambda Beta, and the cell further comprises a nucleic acid encoding Gam and a protein selected from the group consisting of Exo, Arabidopsis F12A21.16, cholera virus orf, B. subtilis YqaJ gene, Listeria phage A118 gp47, B. subtilis phage SPP1 gene 34.1, and African swine virus orf operably linked to the de-repressible promoter.

16. The method of claim 9, wherein the first nucleic acid sequence of interest is single stranded, and wherein the method further comprises introducing a second single-stranded DNA sequence, wherein the second single-stranded DNA sequence comprises more than 10 bp of complementary overlap at either the 5' or 3' terminus of the first nucleic acid of interest.

17. The method of claim 9, wherein the cell is a bacterial cell.

18. The method of claim 9, wherein the cell is a eukaryotic cell.

19. The method of claim 9, wherein the eukaryotic cell is a mammalian cell.

20. The method of claim 19, wherein the eukaryotic cell is a stem cell.

21. The method of claim 9, wherein the de-repressible promoter is the PL promoter.
22. The method of claim 9, wherein the first nucleic acid of interest encodes a selectable marker.
23. A method for inducing homologous recombination in a cell comprising a target nucleic acid, the method comprising:  
contacting the cell with a single stranded DNA molecule and a single stranded DNA binding polypeptide that functions in double strand break repair homologous recombination, wherein the single stranded DNA molecule comprises a sufficient number of nucleotides homologous to the target nucleic acid to enable homologous recombination with the target nucleic acid; and  
introducing the single stranded DNA molecule and the single stranded binding polypeptide into the cell, whereby homologous recombination of the single stranded DNA molecule with the target nucleic acid in the cell is induced.
24. The method of claim 23, wherein the single stranded DNA molecules comprises at least about 70 nucleotides homologous to the target nucleic acid.
25. The method of claim 23, wherein the cell is contacted with a solution that comprises about 0.001  $\mu$ M to about 10 mM single stranded DNA.
26. The method of claim 23, wherein the cell is contacted with a solution that comprises about 0.1  $\mu$ M to about 1 mM single stranded DNA.
27. The method of claim 23, wherein the cell is contacted with a solution that comprises about 1  $\mu$ M to about 100  $\mu$ M single stranded DNA.

28. The method of claim 23, wherein the cell is contacted with a solution that comprises about 0.01  $\mu\text{M}$  to about 10 mM single stranded DNA binding polypeptide.

29. The method of 23, wherein the cell is contacted with a solution that comprises about 0.1  $\mu\text{M}$  to 1 mM single stranded DNA binding polypeptide.

30. The method of 23, wherein the cell is contacted with a solution that comprises about 1  $\mu\text{M}$  to 100  $\mu\text{M}$  single stranded DNA binding polypeptide.

31. The method of claim 23, wherein the ssDNA binding polypeptide is selected from the group consisting of lambda Beta, RecT, P22 Erf, and Rad52.

32. The method of claim 23, wherein single stranded DNA comprises about 20 nucleotides to about 1,000 nucleotides homologous to the target nucleic acid sequence.

33. A method for inducing homologous recombination in a host cell comprising a target nucleic acid, comprising:

introducing into the cell a double stranded nucleic acid comprising a homology arm of sufficient length to induce homologous recombination, wherein the cell comprises a nucleic acid encoding Beta operably linked to a de-repressible promoter; inducing repression of Beta to promote homologous recombination between the double stranded nucleic acid and the target nucleic acid in the host cell.

34. The method of claim 33, wherein the double stranded nucleic acid comprises a 5' overhang.

35. The method of claim 33, wherein the nucleic acid sequence encoding Beta, or a functional variant or fragment thereof, is intrachromosomal.

36. The method of claim 33, wherein the nucleic acid sequence encoding Beta, or a functional variant or fragment thereof, is extrachromosomal.

37. The method of claim 33, wherein the cell further comprises a nucleic acid encoding gam operably linked to the de-repressible promoter.

38. The method of claim 33, wherein the cell further comprises a nucleic acid encoding exo operably linked to the de-repressible promoter.

39. The method of claim 37, wherein the cell further comprises a nucleic acid encoding gam operably linked to the de-repressible promoter.

40. The method of claim 33, wherein the host cell is a bacterial cell.

41. The method of claim 33, wherein the bacterial cell is *E. coli*.

42. The method of claim 33, wherein the *E. coli* is a host recombination defective strain.

43. The method of claim 42, wherein the host recombination defective strain is a rec A<sup>-</sup> strain or a rec A<sup>+</sup> strain.

44. The method of claim 33, further comprising treating the host cell to enhance macromolecular uptake.

45. The method of claim 44, wherein the method to enhance macromolecular uptake is selected from the group consisting of: electroporation, calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, microinjection, liposome fusion, lipofection, protoplast fusion, inactivated adenovirus-mediated transfer, HVJ-liposome mediated transfer, and biolistics.

46. The method of claim 33, wherein the double stranded nucleic acid comprises a nucleic acid sequence encoding a positive selectable marker or a negative selectable marker.

47. The method of claim 33, wherein the homology arm comprises at least 20 nucleotides homologous to the target nucleic acid sequence.

48. The method of claim 47, wherein the homology arm comprises at least 30 nucleotides homologous to the target nucleic acid sequence.

49. A method for cloning a nucleic acid molecule in a cell comprising:  
introducing into a host cell comprising a nucleic acid sequence encoding Exo, Beta and Gam, or a functional variant or homologue thereof, operably linked to a de-repressible promoter, a vector comprising a sufficient number of nucleotides homologous to a target nucleic acid to achieve homologous recombination;  
activating the de-repressible promoter in the host cell, thereby inducing the expression of Exo, Beta and Gam, or the functional variant or homologue thereof  
selecting a host cell in which homologous recombination has occurred between the vector and a target nucleic acid sequence, thereby cloning the nucleic acid molecule.

50. A bacterial cell comprising a defective lambda prophage of genotype  $\lambda$ cI857  $\Delta$ (cro-bioA).

51. The bacterial cell of claim 47, further comprising a nucleic acid sequence encoding a selectable marker upstream of cI857.

52. The bacterial cell of claim 51, wherein the nucleic encoding a selectable marker encodes a protein that confers resistance of the bacterial cell to an antibiotic.

53. The bacterial cell of claim 50, further comprising an inducible promoter upstream of *cI857*, and a nucleic acid sequence encoding a site specific recombinase operably linked to the inducible promoter

54. The bacterial cell of claim 53, wherein the recombinase is Flp, Flpe, Cre, a functional fragment or variant of Flp, a functional fragments or variant of Flpe or a functional fragment or variant of Cre.

55. The bacterial cell of claim 54, wherein the recombinase is Flp, Flpe, or Cre.

56. The bacterial cell of claim 47, further comprising a bacterial artificial chromosome.

57. The bacterial cell of claim 56, wherein the bacterial artificial chromosome comprises a nucleic acid sequence encoding a selectable marker.

58. The bacterial cell of claim 57, wherein the nucleic acid sequence encoding the selectable marker is flanked by recombination sites recognized by the site specific recombinase.

59. The bacterial cell of claim 58, wherein the recombination sites are *loxP* or *FRT* sites.

60. The bacterial cell of claim 56, wherein the bacterial artificial chromosome comprises at least one exon or one intron of a mammalian gene.

61. A method of altering a eukaryotic nucleic acid sequence, comprising:  
introducing a nucleic acid of interest into a bacterial cell comprising an  
extrachromosomal element comprising least one intron or at least one exon of a  
eukaryotic nucleic acid sequence,  
wherein the nucleic acid of interest comprises a sufficient number of nucleotides  
homologous to the eukaryotic nucleic acid sequence to mediate homologous  
recombination,  
and wherein the bacterial cell comprises a de-repressible promoter operably  
linked to a nucleic acid sequence encoding a recombinase; and  
activating expression of the recombinase from the de-repressible promoter;  
wherein expression of the recombinase results in the nucleic acid of interest  
undergoing homologous recombination with the eukaryotic nucleic acid sequence.

62. The method of claim 61, wherein the recombinase is a single-stranded DNA  
binding protein.

63. The method of claim 61, wherein the recombinase is a double-strand break  
repair recombinase.

64. The method of claim 61, wherein the DNA encoding the recombinase  
encodes lambda Beta.

65. The method of claim 64, wherein the DNA encoding the expressed  
recombinase further encodes lambda Exo.

66. The method of claim 64, wherein the DNA encoding the expressed  
recombinase further encodes lambda Gam.

67. The method of claim 61, wherein the extrachromosomal element is a bacterial artificial chromosome, a yeast artificial chromosome, a P1 artificial chromosome, a plasmid, or a cosmid.

68. The method of claim 61, wherein the eukaryotic nucleic acid sequence is a mammalian nucleic acid sequence.

69. The method of claim 61, wherein the nucleic acid of interest encodes an epitope tag.

70. A method of subcloning a DNA sequence, comprising:  
providing the bacterial cell comprising a nucleic acid sequence encoding Exo, Beta and Gam, or a functional variant or homologue thereof, operably linked to a de-repressible promoter, and further comprising a DNA molecule with a target nucleic acid sequence located extrachromosomally;

introducing into the bacterial cell a linear plasmid vector comprising an origin of replication and a selectable marker, and comprising a nucleotide sequence capable of undergoing homologous recombination with the target nucleic acid sequence on the DNA;

inducing the expression of the nucleic acid encoding Exo, Beta and Gam, thereby inducing homologous recombination between the linear plasmid vector and the target nucleic acid sequence; wherein target nucleic acid is inserted into the linear plasmid vector; and

circularizing the linear plasmid vector, thereby subcloning the DNA.

71. The method of claim 70, wherein the DNA comprises at least about 20 kb of the extrachromosomal DNA.

72. The method of claim 70, wherein the DNA comprises at least about 40 kb of the extrachromosomal DNA.



73. The method of claim 70, wherein the DNA comprises at least about 80 kb of the extrachromosomal DNA.

74. The method of claim 70, wherein the extrachromosomal DNA is a bacterial artificial chromosome.

75. A mini-lambda comprising  
a lambda genome comprising a de-repressible promoter operably linked to a nucleic acid sequence encoding Beta, Exo, and Gam a nucleic acid sequence encoding a selectable marker, an attP site and a nucleic acid sequence encoding an integrase, wherein the lambda genome lacks a replication origin.

76. The mini lambda of claim 75, wherein the de-repressible promoter is PL.

77. The mini-lambda of claim 75, further comprising a nucleic acid sequence encoding a selectable marker.

78. The mini-lambda of claim 75, wherein the selectable marker confers antibiotic resistance or antibiotic sensitivity to a bacterial cell harboring the mini-lambda.

79. The mini-lambda of claim 75, wherein the mini-lambda comprises a lambda genome comprising a deletion of the genes from *cro* through *ea59*.

80. The mini-lambda of claim 75, further comprising a gene required for cell lysis, such that the mutation in lambda is deficient in inducing lysis of a bacterial cell.

81. A method for producing a bacterial cell containing a chromosome comprising a de-repressible promoter operably linked to a nucleic acid sequence encoding lambda Bet, Gam, and Exo, comprising  
transforming a bacterial cell comprising an attB site on a chromosome of the bacterial with the mini-lambda of claim 75,  
thereby integrating the mini-lambda into the chromosome at the attB site, and  
producing the bacterial cell.
82. The method of claim 81, wherein the bacterial cell further comprises a bacterial artificial chromosome (BAC).
83. The method of claim 81, wherein the bacterial cell is an *E. coli* cell.
84. The method of claim 81, wherein the *E. coli* cell is a DH10B cell.

FIG. 1  
 nadA gal attL int xis ea8.5 ea22 hin exo bet gam kil clll ssb ral sieB N rexB rexA cl857 (cro-attR-bioA)  
 PL PR Δ

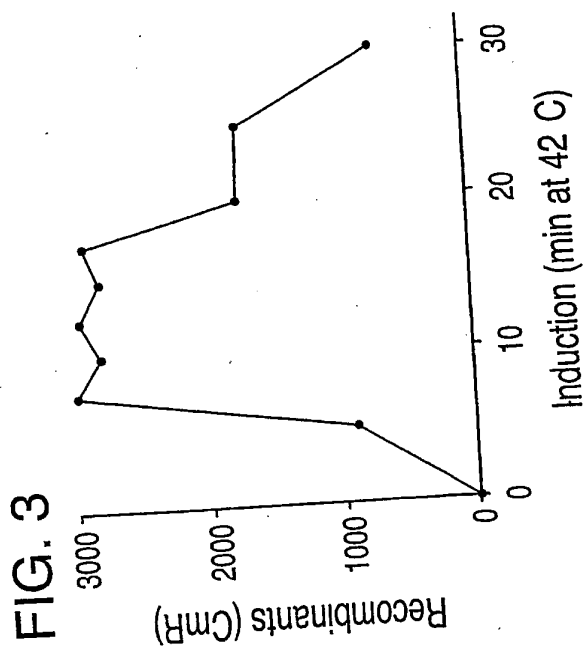


FIG. 2



FIG. 4

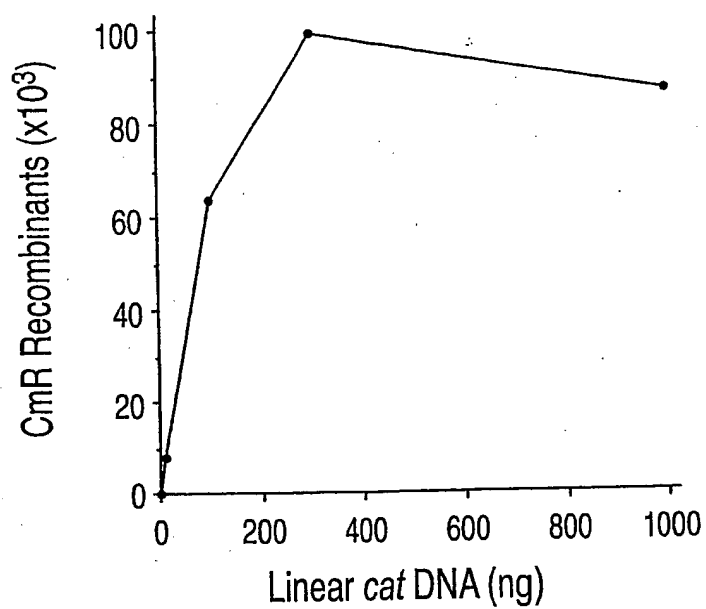
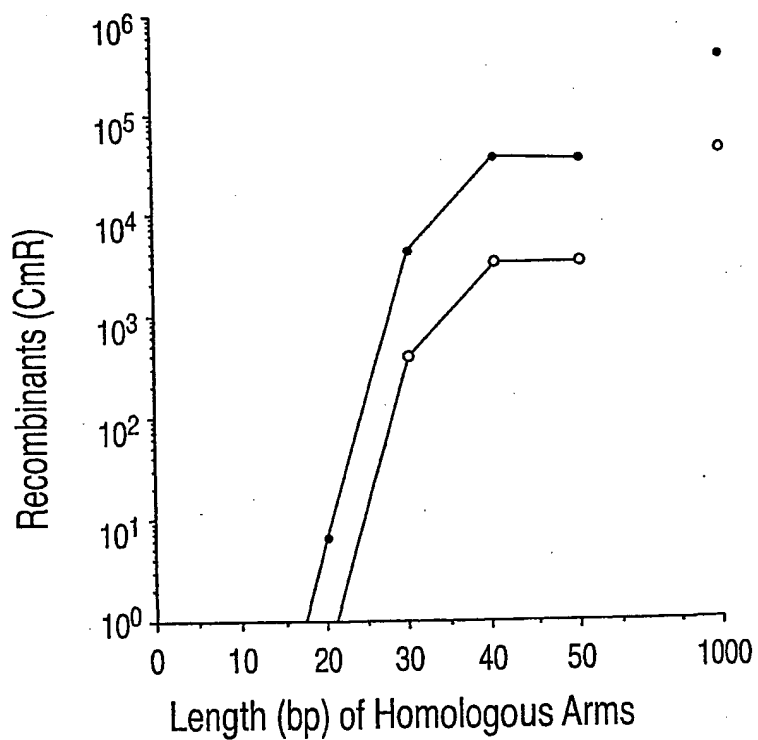


FIG. 5



4/9

FIG. 6

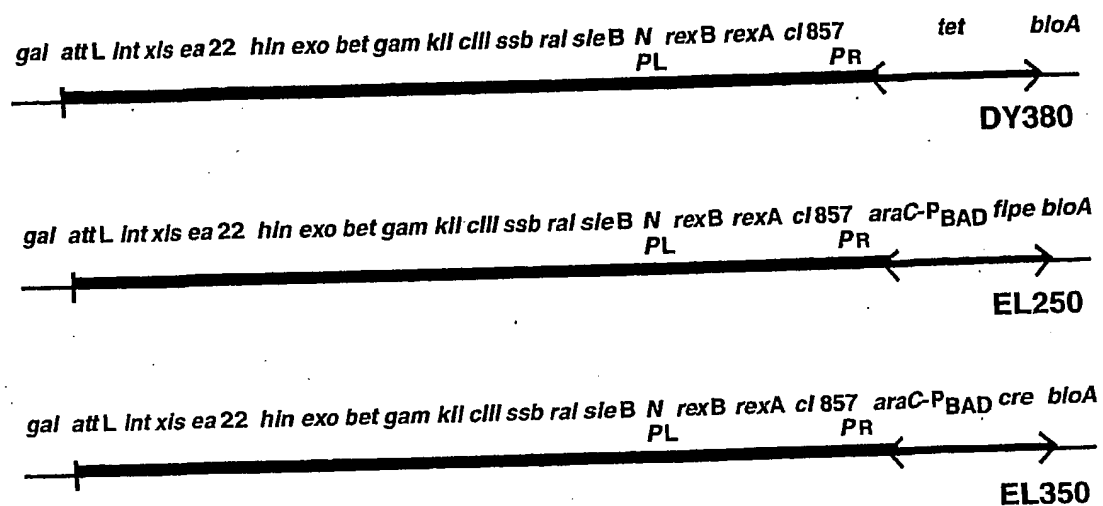


FIG. 7

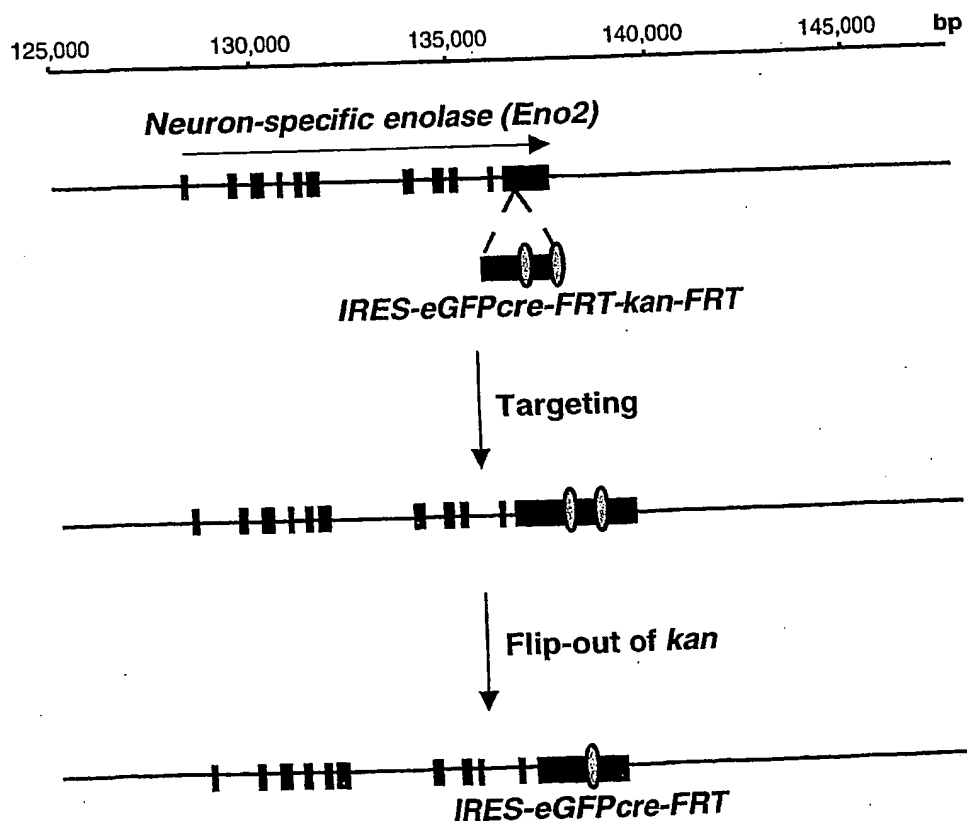
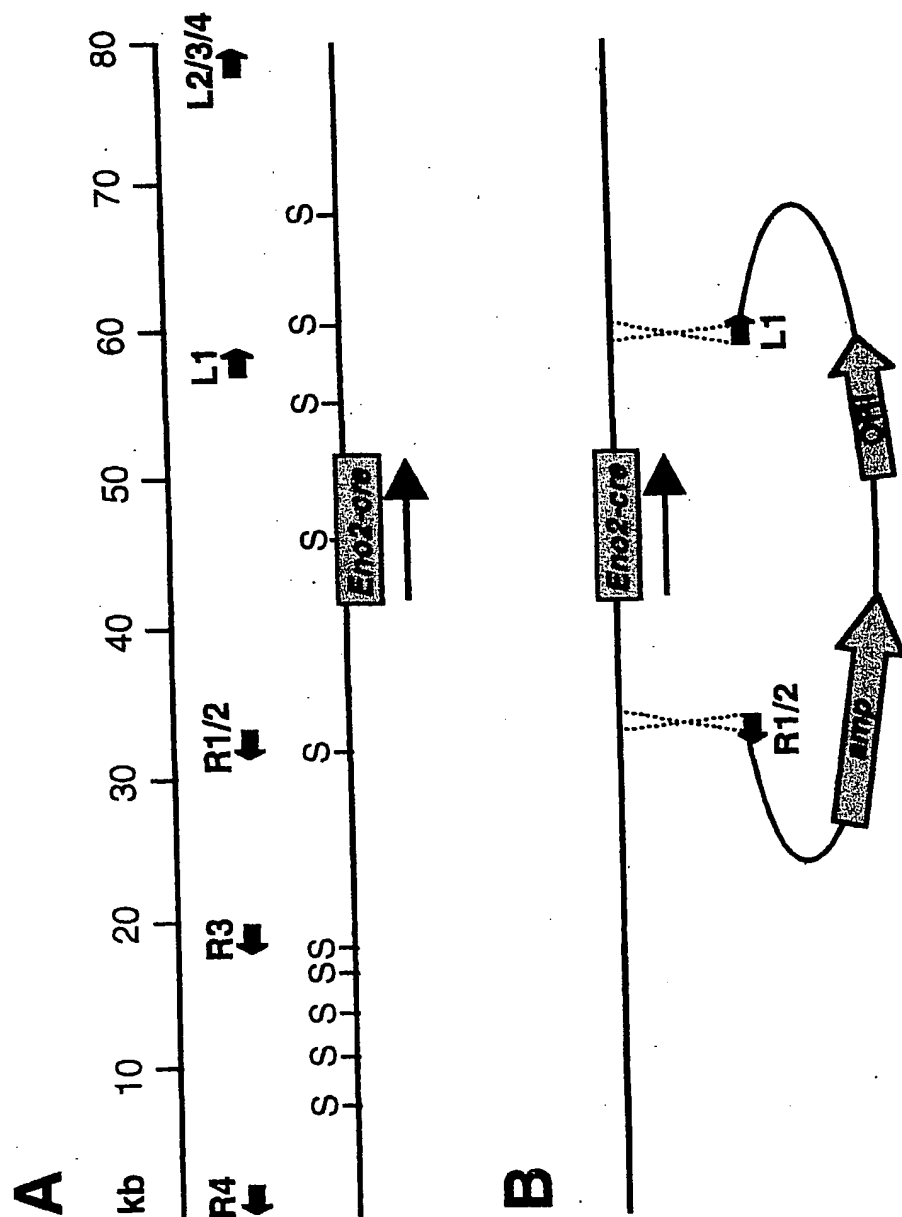
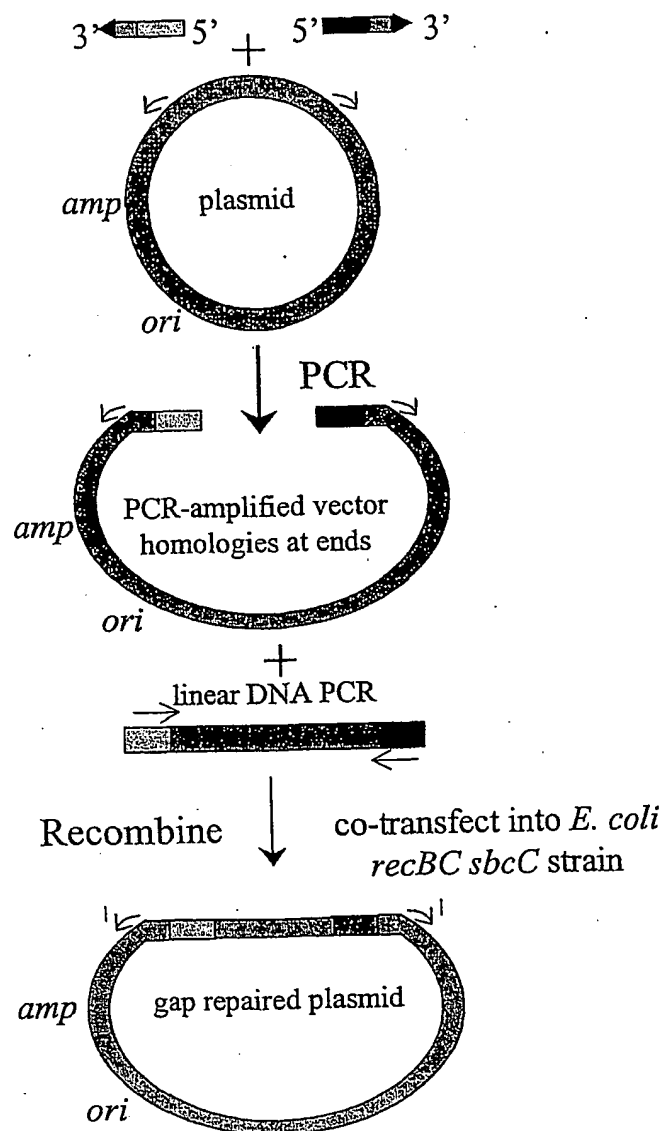


FIG. 8



6/9

FIG. 9





7/9

FIG. 10

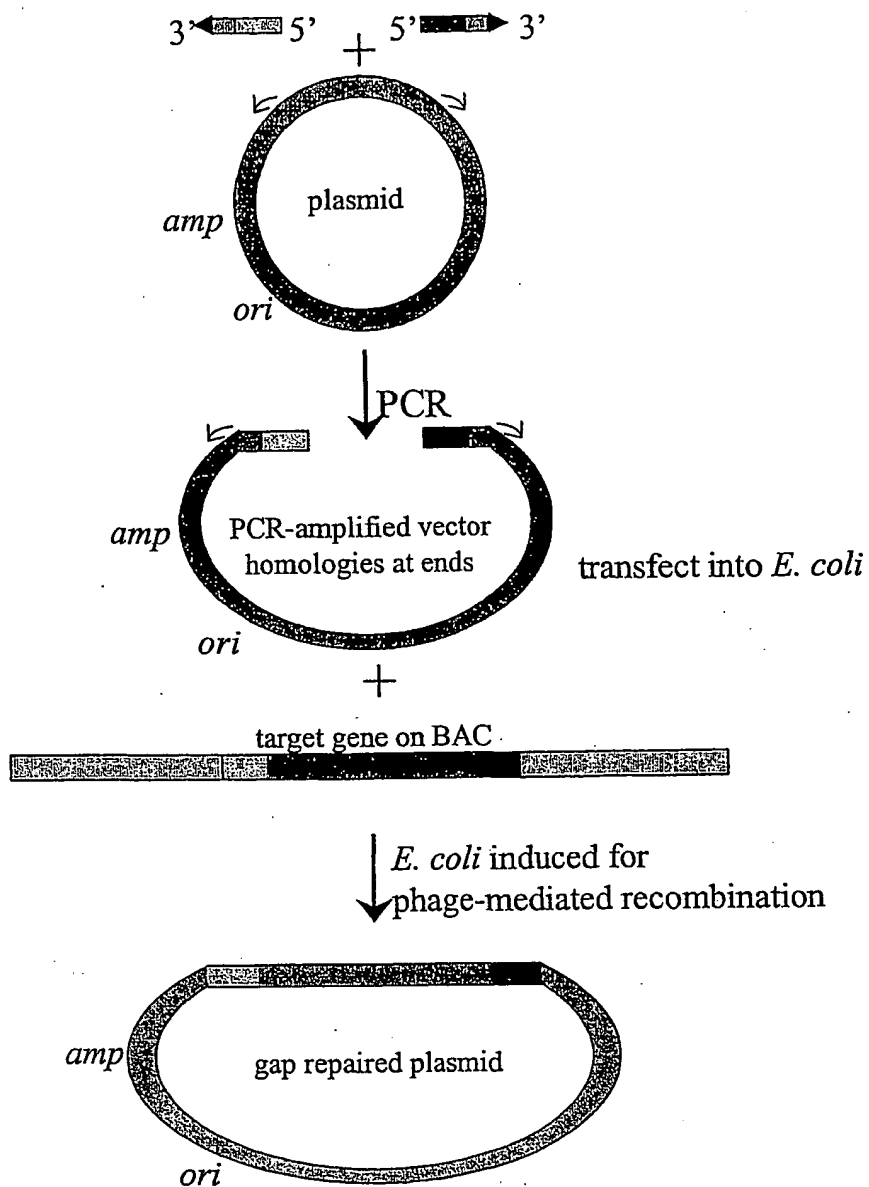


FIG. 11

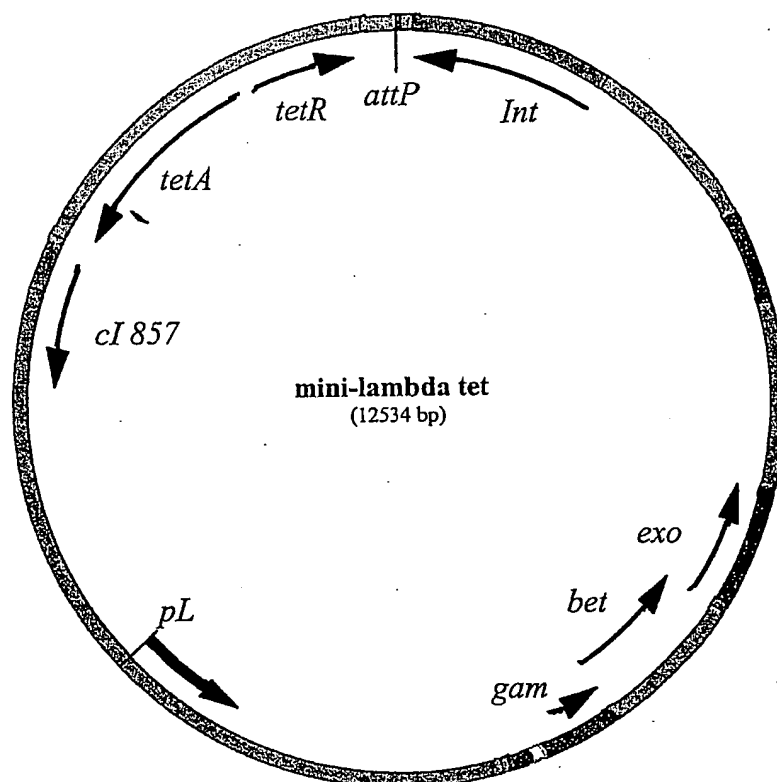
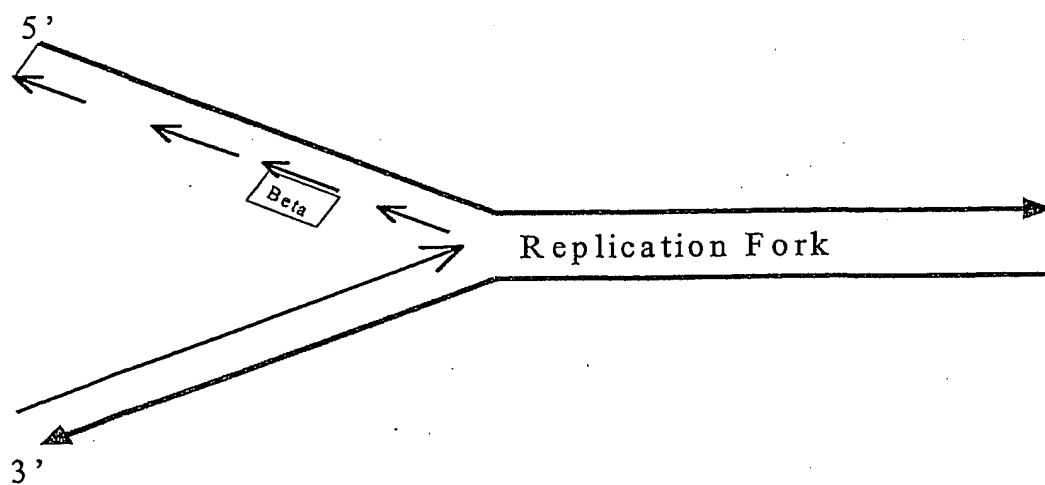
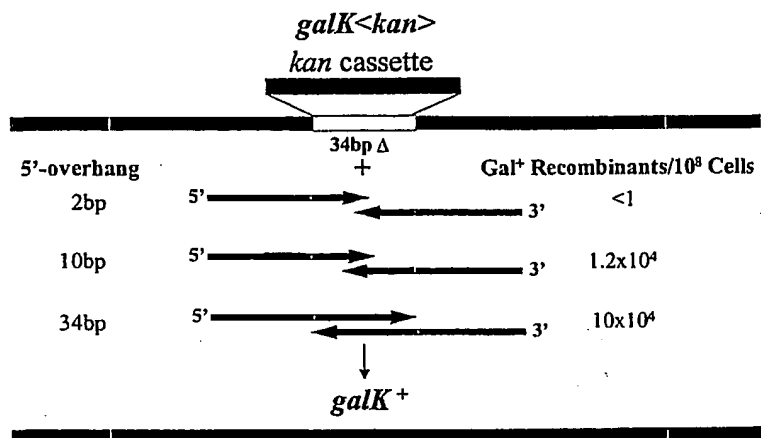


FIG. 12



9/9

## FIG. 13

**Recombination with 5'-overhangs  
The Effect of Overlap Length**

## SEQUENCE LISTING

<110> THE GOVERNMENT OF THE UNITED STATES OF AMERICA, AS REPRESENTED BY  
THE SECRETARY OF THE DEPARTMENT OF HEALTH AND HUMAN SERVICES, NATIONAL  
INSTITUTES OF HEALTH

LEE, E-CHIANG  
COPELAND, NEAL G.  
JENKINS, NANCY A.  
COURT, DONALD L.  
YU, DAIGUAN  
ELLIS, HILARY M.

<120> ENHANCED HOMOLOGOUS RECOMBINATION MEDIATED BY LAMBDA  
RECOMBINATION PROTEINS

<130> 4239-59371

<150> US 60/271,632

<151> 2001-02-26

<150> US 60/225,164

<151> 2000-08-14

<160> 37

<170> PatentIn version 3.1

<210> 1

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 1

tgtgacggaa gatcacttcg

20

<210> 2

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 2

accagcaata gacataagcg

20

<210> 3

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 3

ctcttgggtt atcaagaggg

20

<210> 4  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 4  
actcgacatc ttggttaccg

20

<210> 5  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 5  
cattcaaata tgtatccgct c

21

<210> 6  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 6  
agagttggta gctcttgatc

20

<210> 7  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 7  
tatggacagc aagcgaaccg

20

<210> 8  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 8  
tcagaagaac tcgtcaagaa g

21

<210> 9  
<211> 50  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Homology segment

<400> 9  
gtttgcgcgc agtcagcgat atccattttc gcgaatccgg agtgtaagaa 50

<210> 10  
<211> 50  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Homology segment

<400> 10  
ttcatattgt tcagcgacag cttgctgtac ggcaggcacc agctcttccg 50

<210> 11  
<211> 70  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Homology segment

<400> 11  
aagtcgcggc cggaaccgta ttgcagcagc tttatcatct gccgctggac ggcgcacaaa 60  
tcgcgcttaa 70

<210> 12  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Homology segment

<400> 12  
aacagacacc atggtgcacc tgactcctga ggagaagtct gccgttactg ccctgtgggg 60

<210> 13  
<211> 56  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 13  
ctgcaaggcg attaagttgg gtaacgccag gggtttcgtc aggtggcact ttcggg 56

<210> 14  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 14  
ctcagaagaa ctcgtcaaga agg

23

<210> 15  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 15  
gtaggtacct cgagaatcgc catcttccag caggc

35

<210> 16  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 16  
tcgaattttc tgcaccaat ttactgaccg tacacc

36

<210> 17  
<211> 67  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 17  
gcaagtgtgt cgctgtcgac gagctcgca gctcggacat gaggttgtct tagacgtcag  
gtggcac

60

67

<210> 18  
<211> 69  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 18

catagttaag ccagccccga caccgcgcaa caccgcgtga cgcgaaacctc acgttaaggg 60  
atatttggtc 69

<210> 19  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 19  
gcaggatcca gtttgctcct ggagcgaca 29

<210> 20  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 20  
tgcaggtcga ctctagagga tc 22

<210> 21  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 21  
tggcggtgat aatgggttgc tgtactaagg aggttgatg ctcttgggtt atcaagaggg 60

<210> 22  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 22  
ggcgctgcaa aaattctttg tcgaacaggg tgtctggatc actcgacatc ttggttaccg 60

<210> 23  
<211> 61  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer



<400> 23  
tggcgggtgat aatgggttgca tgtactaagg aggttgatg ctgtgacgga agatcacttc 60

g 61

<210> 24  
<211> 61  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 24  
ggcgctgcaa aaattctttg tcgaacaggg tgtctggatc ctgaggttct tatggctctt 60

g 61

<210> 25  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 25  
tggcgggtgat aatgggttgca tgtactaagg aggttgatg aagcggcatg cataatgtgc 60

<210> 26  
<211> 65  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 26  
ggcgctgcaa aaattctttg tcgaacaggg tgtctggatc ctgtgtccta ctcaggagag 60

cgttc 65

<210> 27  
<211> 63  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 27  
cgcttcgcgg gacataatct cggaaatccc agtgtgctgt gagccaagct atcgaattcc 60

gcc 63

<210> 28

<211> 63  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 28  
gaggctccag gagaatgaga tggtcccgcg ttcaggcaag cgctattcca gaagtagtga 60  
gga 63

<210> 29  
<211> 79  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 29  
gcgagcgtgt gagcgcggt gggcgcccgg caagccgggg ccatggatta caaggatgac 60  
gacgataagg tacaacaga 79

<210> 30  
<211> 79  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 30  
ggccagcaga gcctcagtgt tctccgcgtt gttggtctgt tgtaccttat cgtcgtcatc 60  
cttghtaatcc atggccccc 79

<210> 31  
<211> 80  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 31  
ctctccatgc ctgtctgggt gaggggtggcc cagggggcgat ggctatgaga gaggtcgact 60  
tcttagacgt caggtggcac 80

<210> 32  
<211> 79  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Primer

<400> 32

gcaatgcaga gaagccttgt actgggatga cagagacgga ggggaagagg aggcggccgc 60

gatacgcgag cgaacgtga 79

<210> 33

<211> 80

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 33

gacttctatg acctgtacgg aggggagaag tttgcgacgt gacagagctg gtcgtcgact 60

tcttagacgt caggtggcac 80

<210> 34

<211> 81

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 34

gccccataca cgtaaagtga catagaatca cacagcatca cttctatgga tgcggcgcc 60

gcgatacgcg agcgaacgtg a 81

<210> 35

<211> 79

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 35

catccagtag aacttgggag tgaagctaga gccaaaggcca tctaagtgac aggcggccgc 60

gatacgcgag cgaacgtga 79

<210> 36

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 36

ctgctggaag atggcgattc tcg 23

<210> 37  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 37  
aacagcagga gcggtgagtc

20

STATEMENT ACCOMPANYING SEQUENCE LISTING

The sequence listing does not include matter that goes beyond the disclosure in the international application.

The printout of the attached Sequence Listing is identical to the computer readable sequence listing on the enclosed computer disk.